DOCUMENT RESUME

ED 385 601 TM 024 050

AUTHOR	Pearlman, Mari; And Others
TITLE	An Application of Multimedia Software to Standardized Testing in Music.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ET-RR-93-36
PUB DATE	Aug 93
NOTE	76p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Algorithms; Cost Effectiveness; *Educational Technology; Higher Education; *Multimedia Materials; *Music; Pilot Projects; Professional Training; *Scoring; *Standardized Tests; *Test Construction; Test Format; Test Use
IDENTIFIERS	*Graduate Record Examinations

ABSTRACT

This project explored the possible applications of new technology to standardized testing. Because the multimedia approach is particularly promising for music, the Graduate Record Examination (GRE) music test was chosen as the context for this project. The basic system for administering the test and recording the results was written using Multimedia Toolbook (TM) software by Asymetrix and other programs. The test was designed to replicate some parts of the current GRE music test and to introduce aural questions not currently testable in a paper and pencil format, such as recognition of intervals and chords. The developed test was pilot tested twice in 1992 with 27 and 70 test takers, respectively, all music students with varying amounts of training. A scoring algorithm was developed to simulate expert human judgment of a particular response. The difficulties inherent in test development, administration, and scoring make it apparent that standardized tests using multimedia software need much more development before they can be implemented. Significant resources must be devoted to this development, but such tests could be possible within the next 5 to 10 years if the personal and financial commitments are made. Four appendixes contain the computer screens of the experimental test, the test taker questionnaire, a presentation of performance data (13 tables), and 2 tables of validity data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-93-36

ED 385 601

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

AN APPLICATION OF MULTIMEDIA SOFTWARE TO STANDARDIZED TESTING IN MUSIC

Mari Pearlman
Ken Berger
Linda Tyler

Educational Testing Service
Princeton, New Jersey
August 1993

BEST COPY AVAILABLE

AN APPLICATION OF MULTIMEDIA SOFTWARE TO STANDARDIZED TESTING IN MUSIC

Mari Pearlman
Ken Berger
Linda Tyler

Submitted as the final report of the project supported by the Program Research Planning Council under the title, "Evaluation of Practical Applications of Technology to New Modes of Assessment."

Copyright © 1993. Educational Testing Service. All rights reserved.

INTRODUCTION

This project explored the possible applications of new technology to standardized testing. In the past several years, development of multimedia computer software and programming tools has exploded. The multimedia approach, which combines aural and visual capabilities in the software presentation and the response of the user, is particularly intriguing in such fields as music, in which the simultaneous interpretation of and response to visual and auditory signals is the essence of expertise.

We chose to use the GRE Music Test as the context for this project, since its current paper and pencil format includes direct testing of musical concepts and skills in a quasi-multimedia format (there is an audiotape to which questions are linked). Test takers are required to answer questions using musical notation and their answers are evaluated by expert judges. The field of music has generated a substantial amount of literature on the use of computers in music education (not testing) and this project attempted to apply some elements of this work to standardized testing by adapting some items already developed for the GRE Music Test and developing some new item types particularly suited to computer-based testing.

THEORETICAL FRAMEWORK

Although theoretical explorations of the application of computer technologies to testing have been ongoing for some time (Bennett, 1990; Ward, 1982), few practical evaluations of the possibilities and problems of multimedia computer-delivered, and computer-scored tests have yet been done. This project was intended to assess the practicality of developing, delivering, and scoring items appropriate for standardized testing in the framework of currently available computer technology. We originally planned to recruit test takers who had already taken the Graduate Record Examinations Subject Test in Music so that we might make a preliminary

comparison of performance on the same item types in two different media, paper and pencil and computer. However, this matching of test takers proved too difficult and cumbersome to justify the very tentative comparisons that could be made on this basis, so such a comparison is not part of this study. We did, however, design and implement a test that allowed us to test concepts essential to music education--recognition of intervals and chord qualities--that cannot be tested on a nationally normed and standardized test using the current paper and pencil format.

We hoped to learn some principles of item design, software management, and scoring algorithm design and implementation, that would be applicable to other tests that could usefully exploit multimedia technology. We believed that fields in which simulations of real-life problems or tasks represent the optimum testing stimulus would particularly benefit from what we learned.

We also believed that our experience with some of the current GRE Music Test constructed-response questions that require an expert human scorer indicated that adaptation to a computer and scoring by a computer algorithm might in fact allow an equally efficient measurement of the skills with a cost savings in scoring.

THE TECHNOLOGY

Authoring Software

The basic system for administering the test and recording the results was written using Multimedia Toolbook™ by Asymetrix, running under the Windows™ operating system. This software provides tools for painting controls and images on the screen, as well as a scripting

language for writing programs and controlling multimedia devices.

However, Toolbook-based programs do not offer adequate speed or control to accomplish the dictation functions in the test.[1] Therefore, these functions were programmed using Microsoft 'C' language, version 7.0. A dynamic link library (DLL) was created using 'C' to provide functions to display a dictation window containing one or two staves, respond to user input within the window, and return the results. These DLL functions are invoked from within the Toolbook script. Since Toolbook is designed to facilitate this interaction with DLLs, the Toolbook and DLL functions could be seamlessly integrated.

MIDI files of stimulus material were created using Midisoft™ Studio for Windows (version 3.06).[2] This software permits the entering of tempo, pitch, volume, duration, and attack via notation on a musical staff, or directly, as MIDI events.

---

[1] "Dictation" exercises in music pedagogy are those that present the student with an aural stimulus--a melody, a chord progression, or a rhythmic pattern--and require that the student reproduce what is heard by writing the musical notation on music staff paper that corresponds to what is heard. In this instance, the screen presents a facsimile of music staff paper and the stimulus is played over the headphones via the sound board in the computer.

[2] MIDI, or Musical Instrument Digital Interface, is the acronym that describes the transformation of musical events into numbers a computer can process. A MIDI sequencer plays back digitized representations of musical events (notes, pauses, etc.). There are numerous synthesizers capable of producing digitized representations of any musical instrument. For this project, a Roland 2000 electronic keyboard was used to produce the initial digitized piano sounds for the MIDI file for each item type. In the MIDI file, musical sounds are represented by numerical representations of pitch, duration, attack, decay, and other characteristics, if appropriate.

Testing Workstation Specification

Three workstations were used for testing. They are IBM PC-compatible machines using the Intel 80386 CPU chip, running at 25 megahertz clock speed. All machines contained hard disk drives and were upgraded to 4 megabytes of memory. In addition, these machines are equipped with MIDI ports connected to Kurzweil model 1200 synthesizers. Examinees used headphones connected to these synthesizers.

Development Workstation Specification

Development was done on IBM PC-compatible machines using the Intel 80486 CPU, at 33 megahertz clock speed. These machines are equipped with MediaVision Pro Audio Spectrum 16-bit sound boards.

THE TEST

The test was designed to replicate some parts of the current GRE Music Test and to introduce other kinds of questions not currently testable in a paper and pencil format. It consists of four sections, (1) Intervals, (2) Chords, (3) Harmonic Dictation, and (4) Melodic Dictation. Sections 1 and 2 were designed to test directly the test taker's recognition of intervals and chords played for them through the headphones. Each section consists of 24 items. Test takers hear the intervals and click on the correct description from the list that is displayed on the screen. Each interval or chord is played two times. This kind of item is not currently available on the GRE Music Test because the variability of audio reproduction equipment makes playback of intervals and chords from a cassette unacceptably imprecise and thus unfair to test takers.

Sections 3 and 4 replicate the Harmonic and Melodic Dictation sections of the current GRE Music Test. There are two exercises of approximately equal difficulty in each section.[3] In Harmonic Dictation, test takers hear a chord progression and they are required to place notes in the correct positions on the musical staff displayed on the screen so that the soprano and bass lines of the progression are accurately represented. In this section, the notes automatically appear with correct values, and each measure will allow only the appropriate number of notes to be entered, so that rhythmic errors are not possible. In addition, test takers are asked to consider a set of choices that describe the harmonic function of selected chords and click on the correct description. In Melodic Dictation, test takers hear a melody and notate it on the musical staff displayed on the screen. In this section, note values are controlled by the test taker, and rhythmic errors are possible and are factored into the scoring.

The test was designed to require no more than 60 minutes to complete, including instructional practice items for each section. Care was taken to avoid speededness and to standardize the timing of the items as well as the elapsed time allowed for responses.

The most challenging aspect of test development was the management of the interface between the user and the software. Several assumptions guided our work with the interface. First, we assumed a high level of computer literacy in the testing population. We chose to

---

[3] We base our assumption about difficulty on experience with the same types of stimulus material on the GRE Music Test. The harmonic dictation questions in this experimental test are identical in kind to those on the operational GRE Music Test, where such questions (2 per form) have proved to be consistently similar in difficulty. Melodic dictation stimuli for this test were controlled by the same variables we have found to influence difficulty--number of notes, rhythmic complexity, leaps between notes, inclusion of chromatic notes.

stage the pilot test at Westminster Choir College, which enjoys one of the most sophisticated computer music labs in the United States and consequently a highly computer-literate student body. The recent experience of GRE researchers with randomly chosen test takers (some of whom tried to operate the program by running the mouse over the terminal screen) led us to believe that we would find out much more useful information if we began with a certain level of computer sophistication in the test takers.

Second, we believed that to be useful in a standardized test setting, the working of the software needed to be immediately understandable to the test taker. We wanted the flattest learning curve we could possibly achieve, in order to reserve testing time for test questions rather than mechanical instruction and practice. Thus, the interface had to be as simple and intuitive as possible and the instructions about how to make the software work had to be as lucid and brief as possible.

Third, we believed that in a testing situation the reaction time of the software needed to be as fast as possible. Very sophisticated and complex music software is currently available, but it is both very difficult to use and much slower than the human hand with a pencil. We assumed that the desideratum for the software was that it would work as smoothly and as quickly as a person using a pencil on a piece of music staff paper.

A copy of the instruction and test screens for each section and/or item and the content of the aural stimulus for each (which is the correct response[s]) is presented in Appendix A.

THE EXPERIMENT

The test software was pilot-tested two times, in June 1992 and in November 1992. Both pilot testing sessions took place at the Westminster Choir College Music Computing Laboratory. The first pilot group consisted of 27 test takers, the second of 70 test takers. Test takers were volunteers who were paid $25 for their participation. All test takers were music students, though they varied widely in the amount of training they had completed. With one exception, all test takers were familiar with computers, with music instruction software, and with the use of a mouse and electronic [piano] keyboard in conjunction with music instruction software.

We planned the first pilot test as an information gathering session, using it to evaluate the accuracy of our estimations about timing and difficulty and the adequacy of our assumptions in adapting these item types to computer delivery and response. Test takers were given no verbal instructions at the initial pilot test sessions, other than to put on their headphones and click on START. We intended the test to be entirely self-guiding. At the end of each test taker's session we asked each test taker to complete a questionnaire (see Appendix B).

Reactions to the initial version of the test indicated that our estimation of timing and difficulty for the first two sections, intervals and chords, were quite accurate. Direct aural testing of this kind is not currently feasible on the paper and pencil test; its trouble-free administration on the experimental test was, thus, reassuring. However, we broke no new ground with this item type, since all test takers were required to do was to listen and respond by clicking on the appropriate answer box on the screen. In addition, neither of these item types represented a challenge for computer scoring.

Responses to both kinds of dictation questions, the harmonic dictation and the melodic dictation, indicated that our estimation of timing and difficulty as well as our design for making the software work were fraught with problems for test takers. The universal complaint about all four of the dictation questions was that the tempo of the playings was too fast and that the repetitions were not spaced adequately, though test takers differed on what would constitute adequate spacing. Some wanted more time between repetitions, some less. Test takers also wanted more than four repetitions of the stimulus.

We decided the tempo for the playing of each stimulus on the basis of the tempo used in the paper and pencil GRE Revised Music Test, a tempo that seems to result in acceptable performance levels for the population. We also gauged the number of repetitions of the stimulus--four--and the amount of elapsed time between repetitions of the stimulus on the basis of the paper and pencil test. We added one enhancement to the Melodic Dictation items which is not feasible for a paper and pencil test: we allowed test takers to hear their own notation of the stimulus after the fourth hearing. They were then allowed time to alter their notation before saving and moving on or quitting the test. While we had anticipated that this would make the test easier for test takers and result in better performance, it seemed to make no difference to them. Test takers complained that they could not control the timing of the playback--it was not playable on demand--and furthermore, there was too little time given for notation between playings. The addition of playback as a part of this exercise enhances the assessment of the construct by offering an opportunity for the test taker to compare aurally two musical lines. Given the nature of undergraduate training in aural skills, which concentrates attention on the development of an aural memory, the use of an aural as well as a visual (notation on the screen) cue to the match between the stimulus and the produced

response is an advantage in achieving an accurate reproduction on the musical staff of the stimulus. Thus, we believe that this addition does not fundamentally change the nature of what is being tested so much as it changes the nature of the response possibilities. It is possible to imagine that with playback as an option, much more difficult stimuli might be used to test more sophisticated aural skills than are currently testable in the paper and pencil format.

In addition to the complaints about tempo and timing, test takers were particularly bothered by the mechanics of notation built into the software. The most intractable problem we faced in designing the software was that of relative positions of notes within a measure of music. This is completely straightforward in a paper and pencil setting--one simply "draws" the note wherever it is desired. The exigencies of computer space make this difficult to model on the screen. This difficulty has important implications for these item types, since the physical position of the note from left to right indicates the accuracy of the auditor's memory. On paper, a test taker can indicate by blank space that she/he has missed, say, the first three notes but has heard the fourth and fifth notes. The initial version of the software defined regions for each note in the measure, and clicked the notes into regions as they were entered from left to right. The first click always entered the first note, that is, the left-most note in the measure. Subsequent clicks placed notes one after the other. It was not possible to create a blank space on the musical staff.

The other clear interference with performance occurred in conjunction with the software's mechanism for changing the rhythmic value of a note. There were no note palettes (displays of notes with different values) displayed on the screen because we judged the manipulations

required in the use of such menu bars to be both cumbersome and time-consuming for test takers. Instead, the right button of the mouse was used to change note values, with only those note values needed for these items included in the array. However, the choice was a cycle and the whole cycle often had to be clicked through in order to reach the desired note value. Directions for changing note values were continuously displayed in a box on the screen, but the issue of note positioning was not mentioned in directions.

While both of these features of the user interface--note positioning and changing note values--were indeed cumbersome and in need of improvement, their interference with performance arose from a different and wholly unanticipated source. What caused students to complain was that this test software worked differently from the computer-assisted instruction software with which they were extremely familiar. The test takers were generally unable to adjust to different ways of manipulating the keyboard and the mouse in the time allowed for the test. Certain that we needed computer-literate test takers in order to evaluate our test software without the issue of computer training as a variable, we had not even considered the potential interference that computer literacy itself might create. It seems clear from both pilot testing sessions that the single most difficult challenge in computerized testing requiring sophisticated manipulation of graphics will be the familiarity of the test taking population with software that works differently from the test software. Test takers' persistence in applying the features of the user interface from other software to the test software--of course, unsuccessfully--was quite remarkable.

In addition to the difficulties test takers experienced with the interface, we also experienced intermittent hardware failure. As the "Technology" section explains, the machines we used to

administer the test were equipped with extra RAM to enable them to run the multimedia software. One of the machines had intermittent but persistent problems with the software and simply returned the user to the opening screen without warning. This happened to a total of five test takers during the first pilot test, which explains why there are results for only 23 of the 28 test takers. While we never identified the problem that caused the failure, we did eventually figure out how to minimize the chances of failure by "cold booting" the machine immediately before the program ran. It appeared that the machine would run Windows® in enhanced mode for approximately 55 to 60 minutes before failing. We never determined the cause of this problem, but our makeshift solution eliminated most failures at the second pilot test.

Because the initial pilot test group was very small and also somewhat anomalous in character--these were all summer session students who were not a representative sample of the population--we were reluctant to draw any conclusions from scores on the test. Instead, we decided to alter the software in response to these test takers and try a second pilot.

To address the concerns and complaints voiced by the initial group of test takers, we made several changes in the software and the user interface prior to the second pilot test. All tempos were considerably slowed. In addition, we allowed test takers to control the occurrence of the repetitions of the stimulus, though not the number of such repetitions. The total time allowed for each of the dictation questions was increased, and test takers were told on the initial DIRECTIONS screen the total time, as well as having a time box inset in the screen as they worked.

The ability to position notes within a measure from left to right was also improved. The software was altered so that each measure was divided into "regions", with the number of such areas determined by the maximum number of notes that could be fit into the measure--not by the correct number of notes. When test takers entered a note on the staff the note appeared and remained where it was entered. Other notes could be entered both before and after the first note entered.

The problem test takers experienced changing the rhythmic value of a note was more difficult to solve. We changed the default value to the most commonly occurring rhythmic value in the dictation, but left the rest of the function alone.

We had a much larger and more diverse group of test takers for the second pilot test, and the results of the test are more varied and probably more revealing than those of the first pilot test. It was both interesting and somewhat discouraging to hear the second group of test takers complain almost as consistently as the first about the speed of the dictations. In the case of the melodic dictations a slower tempo would so distort the musical line that the dictation would become more rather than less difficult to remember. While scores on the second pilot test have a much wider range--there are more high scores--most test takers felt the tempos were too fast and that there were too few repetitions.

The alteration of the note placement function reduced the number of complaints about how other computer-assisted instruction software works, but the problem was not really satisfactorily solved. Once entered, notes could not be moved over to make way for other notes--they had to be deleted and reentered. Many test takers noted on the questionnaire

that paper and pencil was much preferable to a computer because of its flexibility in just this area. Furthermore, a number of test takers noted that other kinds of notation outside the musical staff is possible with a paper and pencil test; many test takers use rhythmic notation and note heads off the staff as aides memoire when taking dictation in the traditional fashion.

Test takers commented on the distraction created by the use of the mouse, even if one is computer literate, when trying to listen to and notate a musical line. Most test takers commented on the difficulty created for them by the amount of attention demanded by the medium itself, over and above the test question and its challenges. We believe that this sense of having to pay a lot of attention to the machine and its needs lies behind the complaints about tempo and repetitions.

Several test takers complained in Harmonic Dictation about the necessity to wait for the item to finish all of the repetitions before going forward to the next item. The frequency of occurrence of this complaint was correlated with the achievement of high scores on the test. The software contained no override from one item to the next during the Harmonic Dictation exercises, nor any feature that would allow a candidate to speed up the time between the repetition and the end of the question. In both the initial and the revised versions of the software, this facility--going on at will--was included in Sections 1 (intervals) and 2 (chords) as well as in Melodic Dictation, but it was not included in Harmonic Dictation. The reason for excluding this facility in Harmonic Dictation was the two-part nature of the exercise, first notation and then analysis. The notation had to be largely completed before the analysis section was begun, because only one more playing was permitted in this section.

The most persistent complaint other than tempo was the mouse clicking required to change the rhythmic value of a note. This issue was not satisfactorily resolved in either version of the software.

Of the 70 test takers, 65 completed the test successfully. Three test takers experienced hardware failure during their test session--one of the three at a machine other than the problem machine identified during the earlier pilot test. Appendix C presents the results of the test in several different graphic presentations. Any interpretation of the score distributions is fraught with potential error since this experiment is laden with confounding variables. In order to bolster these data with concurrent data, we interviewed the aural skills teachers of all of these students, and they shared with us the achieved grades in relevant courses as well as their assessment of the ability of the student to perform these kinds of exercises. This concurrent data appears in Appendix D. These two appendices taken together indicate that there is some correlation between high achievement in class, high ability, and high scores on the sections of our experimental test. It appears to discriminate between students who work hard but have little natural aural ability and those students who may achieve little in class but enjoy great natural ability. This test favors the test taker with bountiful natural ability. There is also some suggestion in these data that high computer literacy compensates for other deficits--like language ability--on this test, and may give its owner an advantage when he or she is compared to a non-computer literate, or less computer literate peer of similar ability.

THE SCORING

A primary motive for this research project was our desire to assess the validity of our judgment that certain of the item types on the paper and pencil test could be computer scored if the program and equipment to read the test takers' responses were available. The dictation questions require an expert judge because of the form of the responses, but expert judges virtually never disagree on their judgments of a particular response. Thus, we were particularly eager to see if we could devise a scoring algorithm that could simulate expert human judgment on these questions.

We used as the basis for the algorithm rules the rubrics used for harmonic and melodic dictations at the GRE Music Test scoring session. We attempted to replicate the decisions that an expert human scorer would make on each of the dictation questions.

For harmonic dictation, the rules were quite simple. Test takers were not allowed to change the note values or to add extra notes in any measure; thus, rhythm errors or correct pitches displaced from their proper position were not considered in the scoring. The algorithm simply matched the pitch for each note in the treble and then each note in the bass. The boxes containing harmonic analysis beneath the chords were scored separately bv matching the intended Roman numeral with that selected by the test taker.

Melodic dictation presented more difficult issues. While it was simple enough to create a numerical value for each pitch and note duration and then match the given string of pitches and durations with that notated by the test taker, more complex errors occur in melodic dictation. For example, a test taker may skip a note and then pick up, correctly, on the next

five or six notes. However, the computer did not allow for an empty space indicating a note left out, so all pitches notated in the spaces would be counted as errors. To accommodate this possibility, we had the algorithm compare pitches to the key from left to right and then from right to left. Any string of pitches correct from right to left was credited. This scoring rule does not allow, however, for the correct rhythmic placement of a pitch in a measure, even when that pitch is surrounded by other wrong notes. When, according to human judgment, a test taker notated only one or two pitches correctly in a measure--that is, the correct pitch on the correct beat--if these pitches were intermixed with other wrong pitches, they were not credited.

Two other errors for which we had no provision in the scoring algorithm also occurred in a number of cases. The first is the problem of added pitches. If the test taker notated too many notes in a measure, the algorithm simply compared the pitches of the number that should have been there and ignored the additional pitches. A test taker might, thus, get credit for the first three notes in a measure because only three notes belong in that measure and have no credit taken off for the fourth and fifth note added in. In addition, a test taker might interpolate two or three pitches in the middle of a measure and then get back on track by the third and fourth beats. If the note values allowed the human judge to say that the third and fourth beats were indeed heard correctly, credit would be given for them and a point subtracted for the added pitches. The algorithm in this case, however, would subtract credit for the number of notes in the measure because it simply matched note by note without regard for note duration and beats. For example, in a dictation in which there are four beats per measure and each beat is a quarter note in duration, two eighth notes can occupy the "beat space" of a single quarter. A test taker might begin the measure with four incorrect

eighth notes, and then notate two correct quarter notes for beats 3 and 4. A human judge would subtract two points for the two wrong pitches, regarding the third and fourth pitches notated as added pitches. Then credit would be given for the correct third and fourth beats. The algorithm was not written to accommodate these situations, which arose with some frequency, in about 16 of the 68 melodic dictation sections completed. The algorithm could accurately score the measures in which added pitches were notated only in cases in which the last "n" pitches of the measure were correct. In this latter case, the backward scanning built into the algorithm sufficed to count these pitches and credit them.

Scoring Algorithm

The scoring algorithm was implemented using Microsoft FORTRAN version 5.1. The basic key information is read in from a response file previously created by running the test delivery system and answering each item correctly. Interval and chord identification were done by simple comparison of the responses to the key. Harmonic dictations were matched note for note against the key, with each correct note worth one point. Harmonic analysis was scored one point per chord.

Melodic dictations were first scored by measure, note for note against the key, in the forward direction, comparing pitch only. A point was subtracted for each pitch in the key which was not matched in the response. Next, if the first note of the measure was correct, and there were more notes in the response than the key, a point is subtracted for the extra notes, and the pitches are compared from the end of the measure back to the beginning. A point is added back for every matching note not already credited in the in forward direction. Octaves and redundant accidentals are considered to be correct.

Next, melodic dictations are scored for rhythm, with one point subtracted for each wrong measure. Any wrong, missing or added note causes the entire measure to be scored as rhythmically incorrect.

CONCLUSIONS

The overarching conclusion to be drawn from this experiment is that standardized tests using multimedia software are not just around the corner unless significant financial and personnel resources are devoted to their development. We could develop an online GRE Music test and have it ready for delivery in 18 months. It is questionable, however, whether any program would be willing to invest the money it would take to accomplish this delivery. And it would be equally problematic to find a venue to deliver it other than our pilot test venue, Westminster Choir College. Even without extraordinary efforts and expenditure, though, such tests are well within the possibilities of the next five to ten years. We have organized our conclusions from the most significant and generally applicable to the most detailed and limited to this study.

Continuous Tryout

It was clear from the development stage of the initial version of the test that the most difficult parts of the test development process are the on-screen directions and the design and management of the interface. Much of the development of directions and the interface is guesswork--how will people respond to this? will this be easier than that? can someone understand what procedure to follow from these written directions? In some ways, the most efficient development method is to arrange for continuous tryout if such a thing were possible. Actually watching people use the software and listening to their questions and

soliciting their complaints is the most revealing and helpful way to get this part of the process right. Presumably this problem will lessen as a greater proportion of the test taking population becomes fully computer-literate and more assumptions about understanding can safely be made. However, it is also possible that once such a basis can be assumed, software developers will build on it and the user will be expected to perform ever more complex feats of manipulation to accomplish the task at hand.

Perhaps the most important consideration in planning for the tryout of any online assessment instrument is how to secure a population that can serve as the locus of a continuous tryout procedure. This might mean that in planning the delivery of the prototype, one cost would be the honoraria over the course of six months to some number of test takers or users, and the project would be timed so that a revised version would be delivered at regular intervals of, say, a month throughout the testing period. This would add to development costs, but handsomely repay this early investment by yielding a product that is virtually ready for production. We believe that had we known to arrange for continuous tryout and the ongoing and frequent software revisions such a plan entails, the prototype music test would be quite sophisticated by now.

In planning for a continuous tryout population, one would want, ideally, half the people to remain stable over the course of the whole development cycle and half to change at each new version, since one of the goals is to create a version of the software that works nicely with users who have not yet seen it. On the other hand, the feedback (better or worse, harder or easier) to be gained from the users who have seen each iteration is also valuable. It is also our intuition that one might learn a great deal from the stable portion of the tryout

population as they became more and more familiar with the goals of the software and the constraints of the medium.

Interference from Other Software

The interference engendered by test takers' familiarity with the conventions of other software in the same field as that being tested is a more intractable problem, if indeed what we found in this experiment is generalizable to other fields. We have nothing but anecdotal evidence for this assertion, but it seemed very clear that performance was adversely affected by test takers' inability to respond to the test software without imposing the structure of other software with which they were already familiar. It is interesting to note, though, that we had only one test taker who had never used a mouse for any task or a computer to do any task associated with music. This person was among the top ten scorers on the test.

The most obvious solution to the problem is to build software that uses the conventions common to already developed software with which users are familiar. However, music software packages developed for computer-assisted instruction and for composition do not have much in common. In fact, it seems a point of honor for software developers to do it differently, so the assessment prototype was based on our own sense of the best way to proceed. We suspect that most fields are not markedly different in this regard from music. Thus, the most obvious solution is probably not the most feasible.

Part of the solution to this problem may be a more extensive tutorial, but the developer must be careful in constructing such a tutorial. It is probably best to be familiar with as much software in the field of the prototype as possible, particularly that software that test takers

might know well. This would allow the tutorial to address specific areas of possible confusion, particularly if there are good reasons for the prototype software to accomplish tasks or functions in a way different from that in commonly used software (as was the case in our experiment). One could, of course, given sufficient funds, incorporate the several common ways of getting things done electronically that are used in widely known software, and allow the test taker to choose his or her favorite. Also, more extensive tutorials have their own disadvantages. Users rush through them, growing tired long before they have mastered all the arcane instructions needed to make using the software itself not part of the test. Big tutorials take lots of time away from the testing time, making the assessment itself more cumbersome.

We believe that the optimal approach--though it is not a solution--would be to examine the most commonly used software packages in the company of the programmer who will be writing the code for the assessment prototype. This kind of scrutiny does yield, eventually, some common threads, however different the software packages appear on the surface, and the programmer can immediately identify the common underpinnings of the elements in the software packages. We did not proceed in this fashion out of ignorance rather than conviction, but it seems clear from our experience that it would have been immensely helpful had we laid the groundwork for our prototype in this fashion.

Hardware Problems

Hardware requirements of sophisticated software like that devised for this experiment will most likely be a problem for some time to come. As the experiences recounted above make clear, intermittent equipment failure is an operating condition, and standardized testing would

have to make some provision for coping with such failure. Perhaps more significant, though, is the gap between machines that are widely available in educational settings and the requirements of sophisticated and complex testing software. For every advance in the power and speed of personal computers, there is a corresponding but unequal (i.e., greater) advance in the programming possibilities for software. And the resulting software packages require very high quality hardware to run adequately. In addition, while the power of the programming tools may make the user interface appear more and more elegant, spare, and friendly, the skill required of the programmer will increase dramatically, since ease-of-use is generally based on more complex coding.

While it is discouraging to be solidly in the middle rather than in the vanguard, it seems clear that software packages that rely for their efficient function on very sophisticated hardware that is unlikely to be generally available in the settings frequented by the testing population will not be successful operationally. Clever programming can compensate for some of the disadvantages of less sophisticated hardware, but at present, most widely disseminated tests would not be well-served by software development that required very high level hardware.

Future Directions & Enhancements

If we were to redo the experimental test, there are several enhancements and changes we would make. First, to address test takers' dissatisfaction with the note value functions in the software, we would experiment with alternatives to setting note values, such as a "floating note palette." This would require test takers to click on and drag the correct note to the proper spot. We might also consider supplementing the mouse interface for setting note values with a keyboard interface.

To counteract the problem of a steep learning curve for the software used in the test, we would try using animation in the instructional material on the notation interface to show how the mouse is manipulated to accomplish specific tasks in notation:

a. Placing notes on the staff initially

b. Altering a pitch

c. Inserting pitches between sets of pitches

d. Deleting pitches

e. Changing note values

Finally, we believe that the use of audio instructions to supplement or replace on-screen directions would be an effective aid in the testing situation. Since the test delivery workstations are equipped with sound boards, why not make use of them?

All of these enhancements would be best explored in continuous tryout, to return to the beginning of the end. Some may not work very well, others may have unanticipated disadvantages, still others may lead to unforeseen spinoffs.

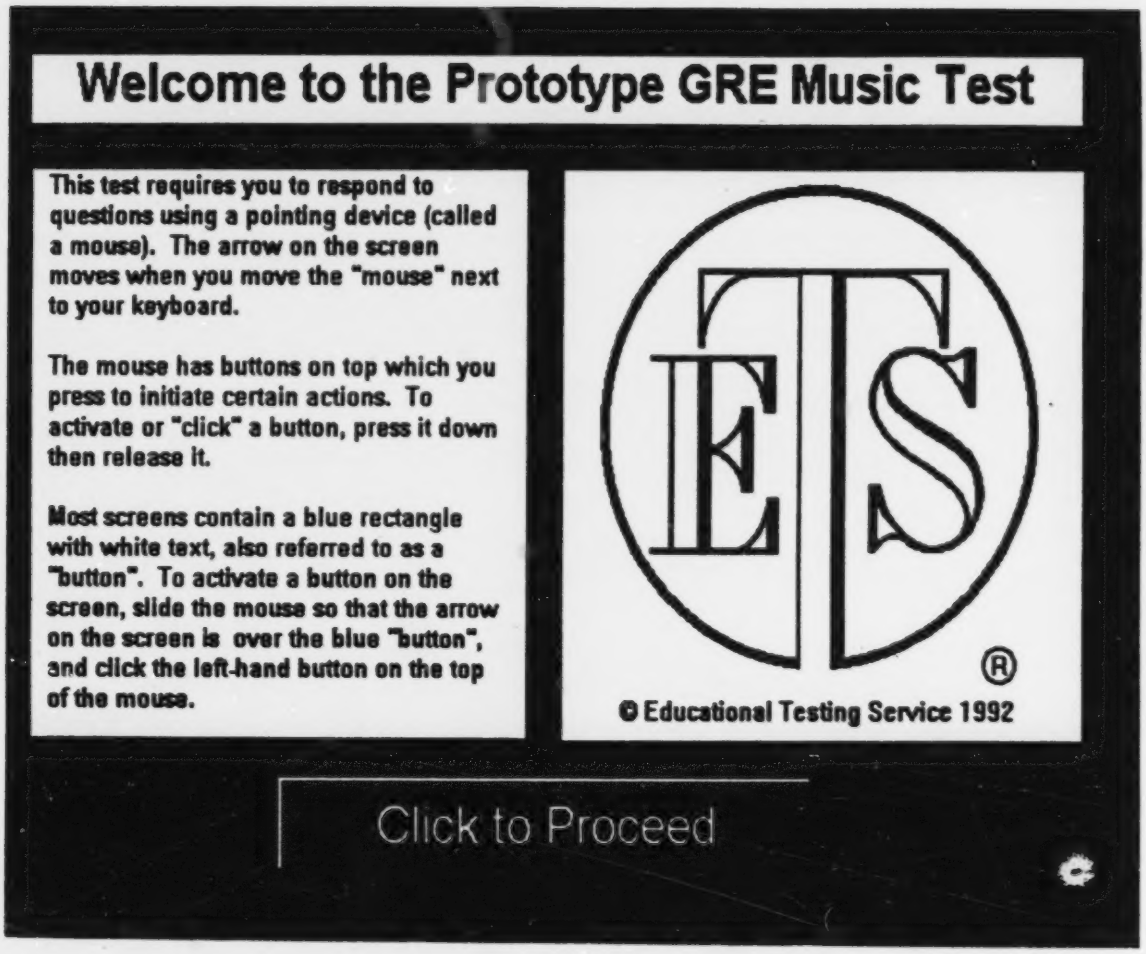
Problems Remaining to be Solved

We have listed here the problems for which no solution is currently available, though given the speed of current development, these may be solved by the time this report is printed.

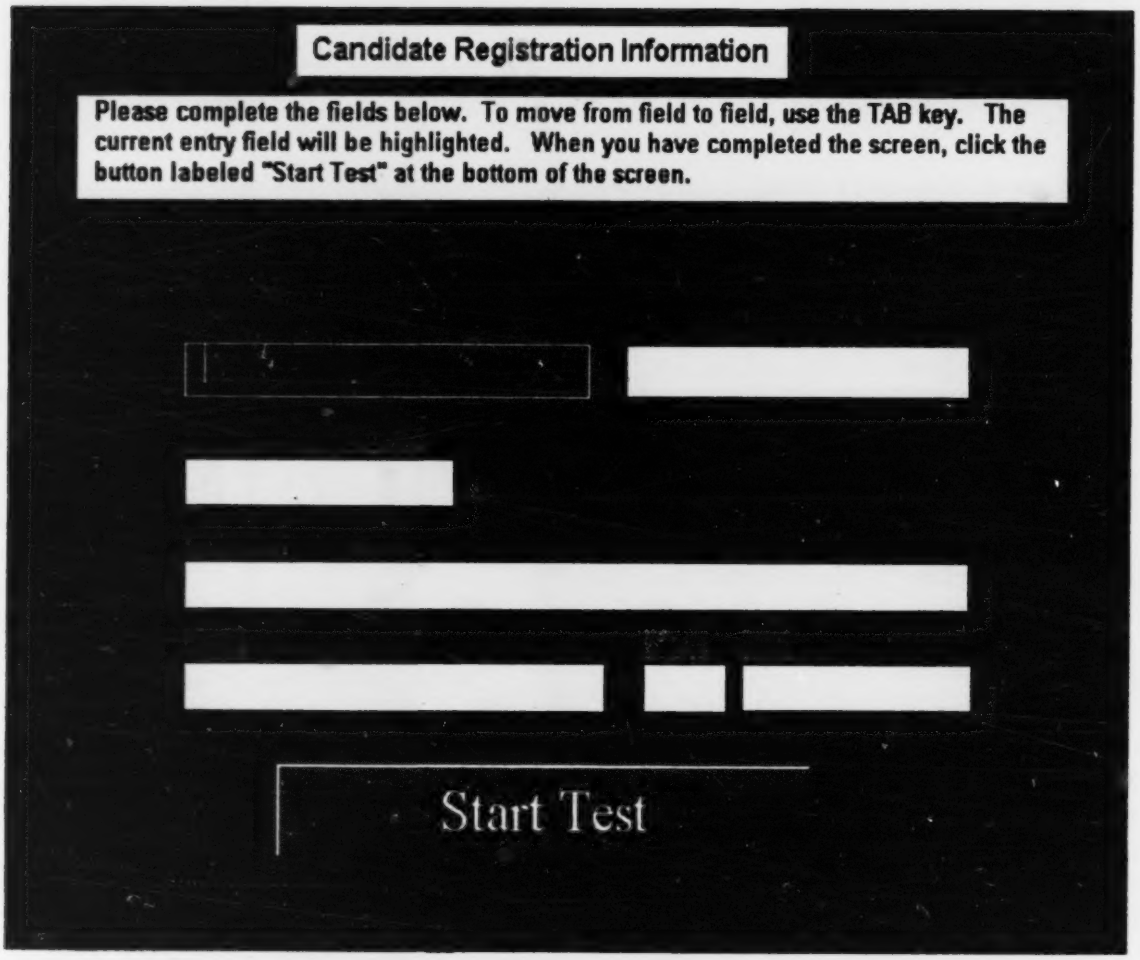
1. Problems arose porting MIDI files from MAC to PC. There are apparently a substantial variety of MIDI formats and very little standardized information about the varied formats.

2. One testing machine had intermittent problems and would occasionally and unpredictably exit Windows, interrupting a testing session and possibly causing the loss of data. No cause for this problem was ever identified and such intermittent and puzzling breakdowns are the bane of new software everywhere. In the testing environment, they are particularly serious.
3. An easily soluble problem, given enough money, is that of the performance on the testing workstations. This was marginal, as would probably always be the case when the test software uses a high-level package such as Toolbook. The software required tuning and optimization for the testing environment. Performance was so much better on the development machines, that it was sometimes difficult to gauge, without actually trying it, whether it would be acceptable on the testing machines. This is another reason for the adoption of continuous tryout, as described in the Conclusions section.

APPENDIX A



The 29 different screens that make up the experimental test are presented below in order of their appearance. The test was intended to be entirely self-guiding. Mouse-driven buttons for proceeding through the test appear at the bottom center of each screen. Captions pointing out features new to a screen appear beneath each facsimile below.


Candidate Registration Information
Please complete the fields below. To move from field to field, use the TAB key. The current entry field will be highlighted. When you have completed the screen, click the button labeled "Start Test" at the bottom of the screen.
Start Test

Directions for Section I – Intervals

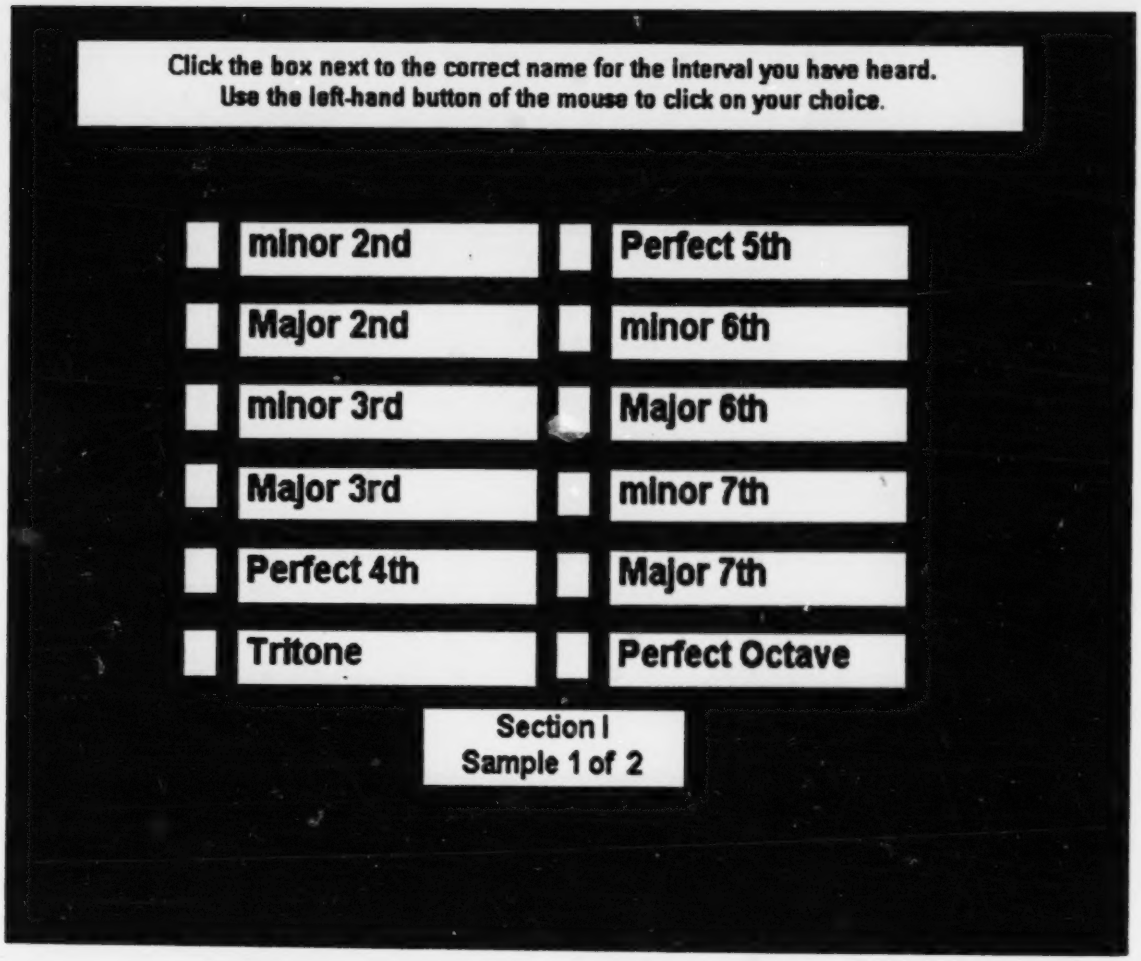
In this section you will hear 12 melodic intervals followed by 12 harmonic intervals. You will hear each interval two times, with a 10-second pause following the first playing and an 8-second pause following the second playing (you will see a warning message when 5 seconds remain). The current question number is displayed at the bottom of the screen.

For each interval, click on the box next to the correct name for that interval. **USE THE LEFT-HAND BUTTON OF THE MOUSE TO CLICK ON YOUR CHOICE**. You may change your choice by clicking on another box. The box that is highlighted when the next interval begins will be stored as your answer.

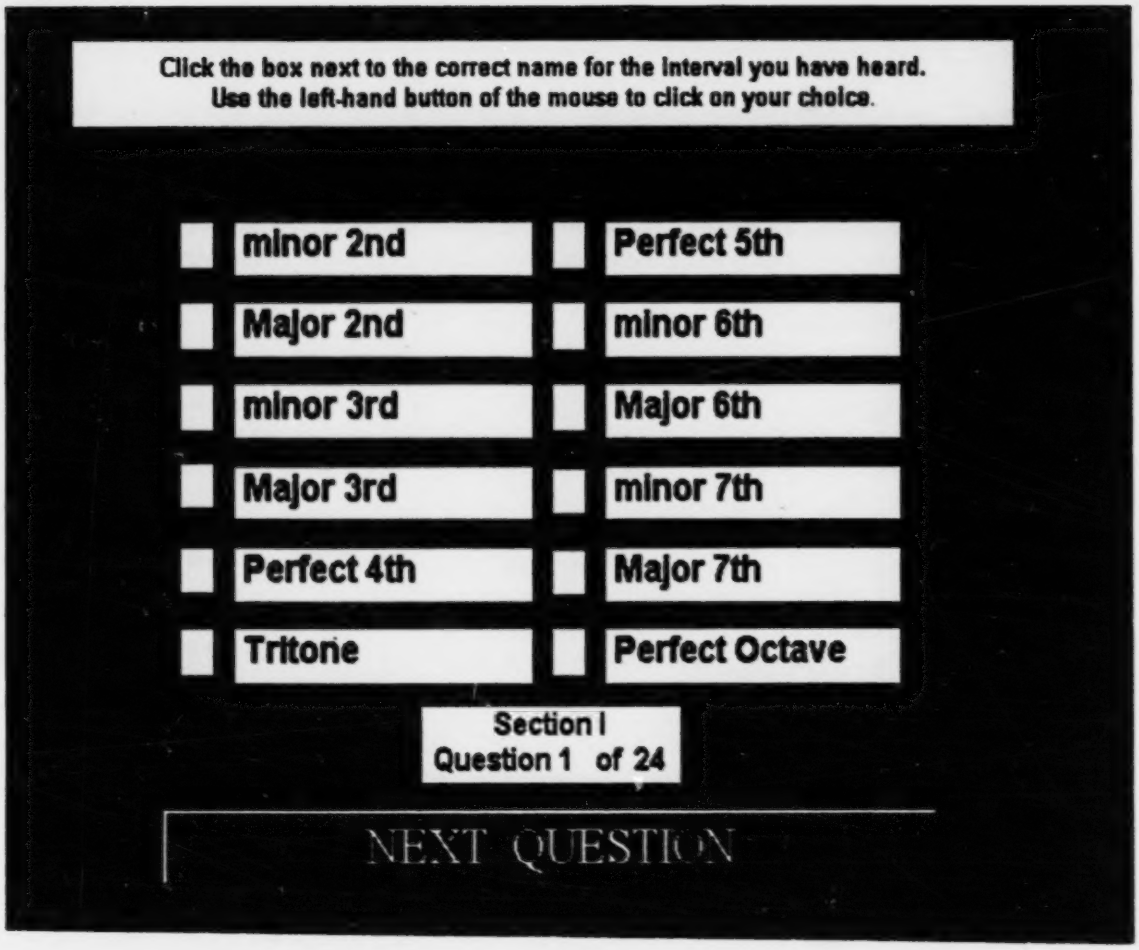
If you do not need all of the allotted time to answer, you may click the button labeled "**Next Question**" at the bottom of the screen. Otherwise, you will proceed through the test at the set pace.

When you have finished reading these directions, click on the button labeled "**Begin Sample Question**" below, so that you may practice answering one sample question for each interval type.

Begin Sample Question



This screen for the sample question is identical to the screen presented for the Section I questions. Candidates hear an interval played on a piano over their headphones and then click on the correct identification of the interval they have heard. There are two sample questions, one melodic interval and one harmonic interval.



This is the screen that appears for all Section I questions. The number at the bottom center of the screen automatically advances when the test taker clicks on NEXT QUESTION.

Directions for Section II – Chords

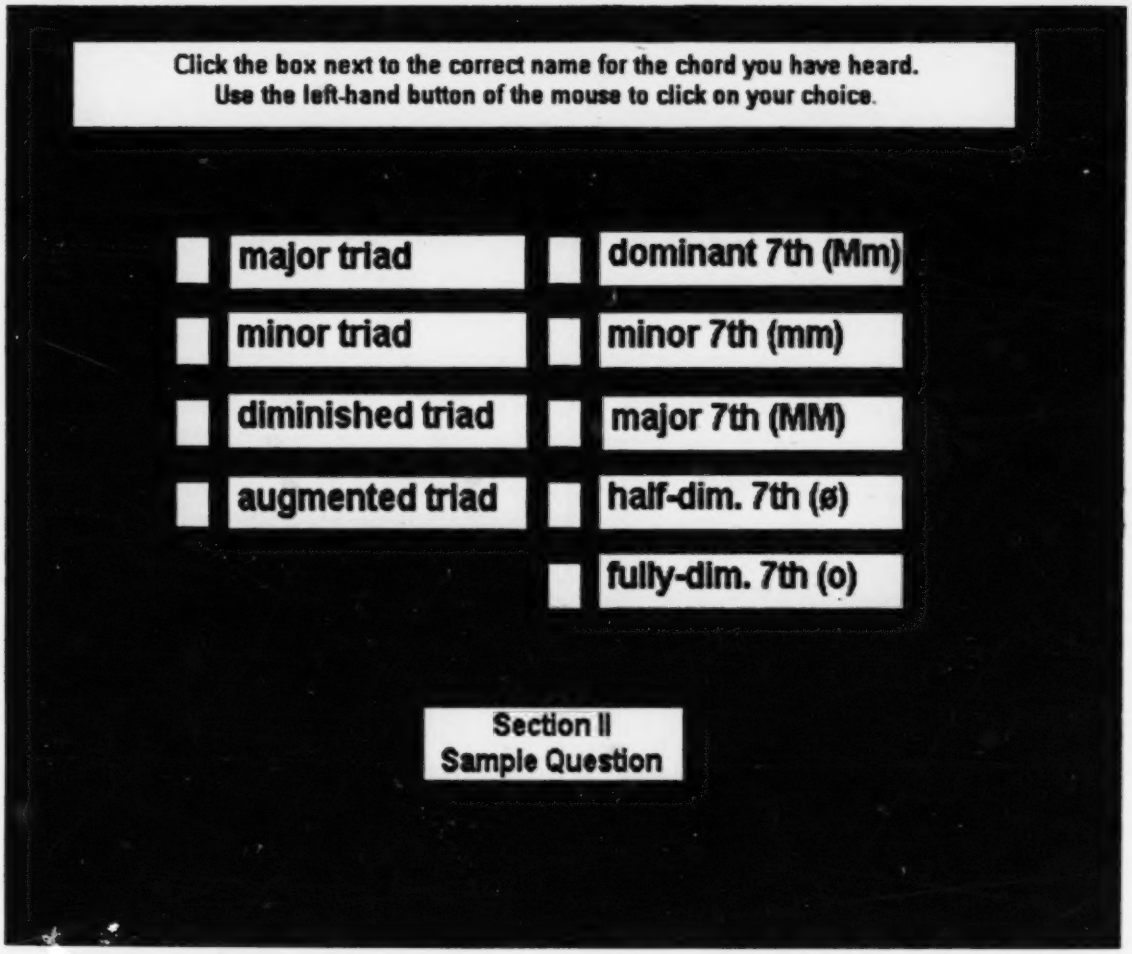
In this section you will hear 24 chords. All of the chords are in root position. Each chord will be preceded by its arpeggiated pitches. You will hear each chord two times, with a 10-second pause following the first playing and an 8-second pause following the second playing (you will see a warning message when 5 seconds remain). The current question number is displayed at the bottom of the screen.

For each chord, click the box next to the correct name for that chord. **USE THE LEFT-HAND BUTTON OF THE MOUSE TO CLICK ON YOUR CHOICE**. You may change your choice by clicking on another box. The box that is highlighted when the next chord begins will be stored as your answer.

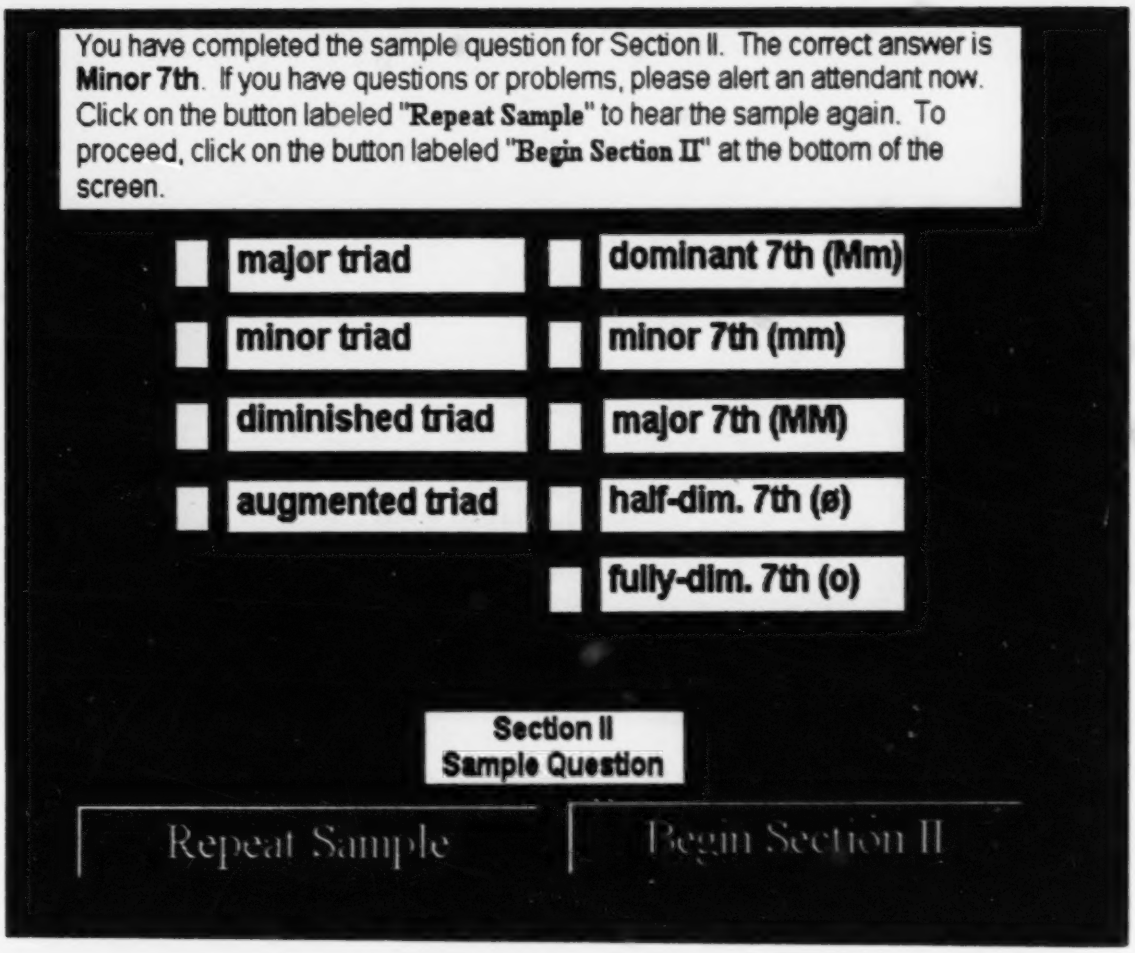
If you do not need all of the allotted time, you may click the button labeled "**Next Question**" at the bottom of the screen. Otherwise, you will proceed through the test at the set pace.

When you have finished reading these directions, click on the bar labeled "**Begin Sample Question**" below, so that you may practice answering one sample question for each interval type.

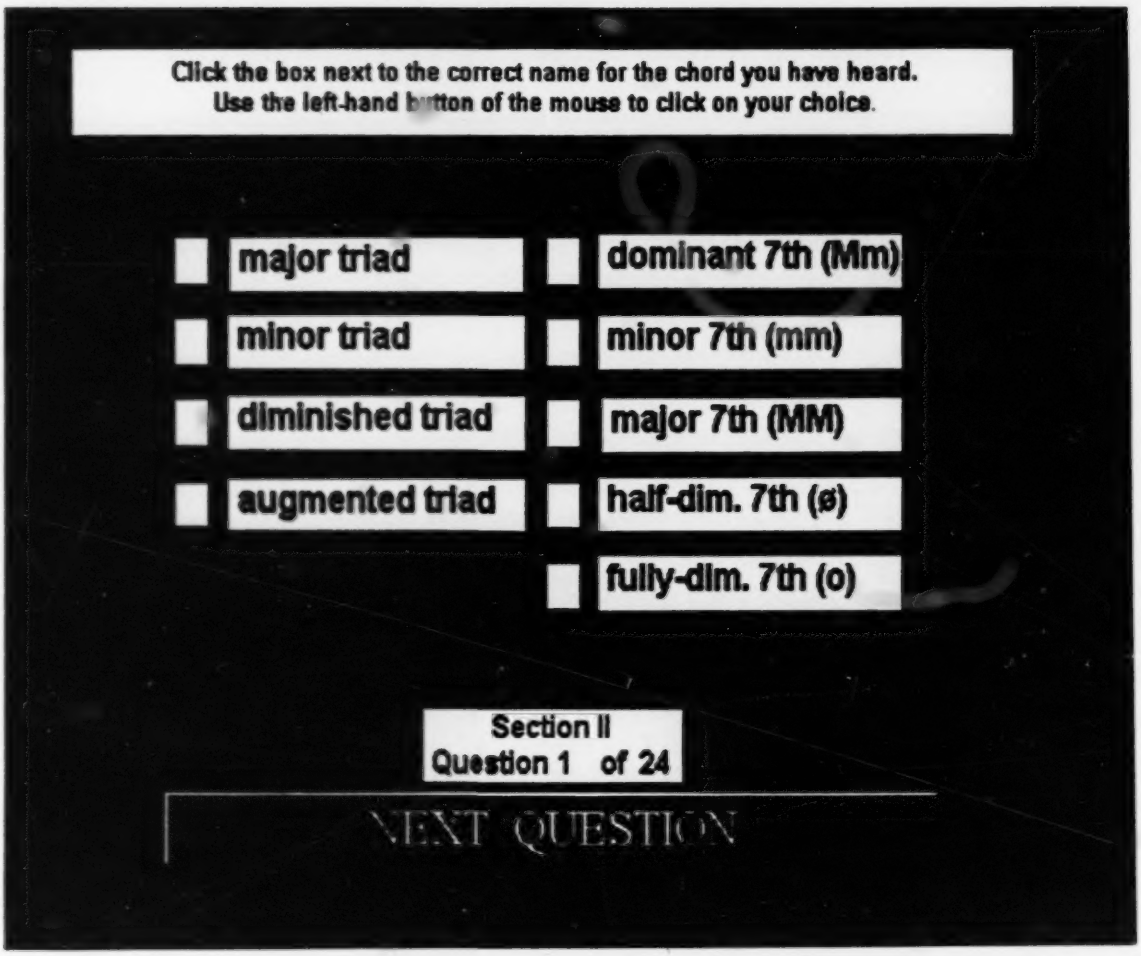
Begin Sample Question

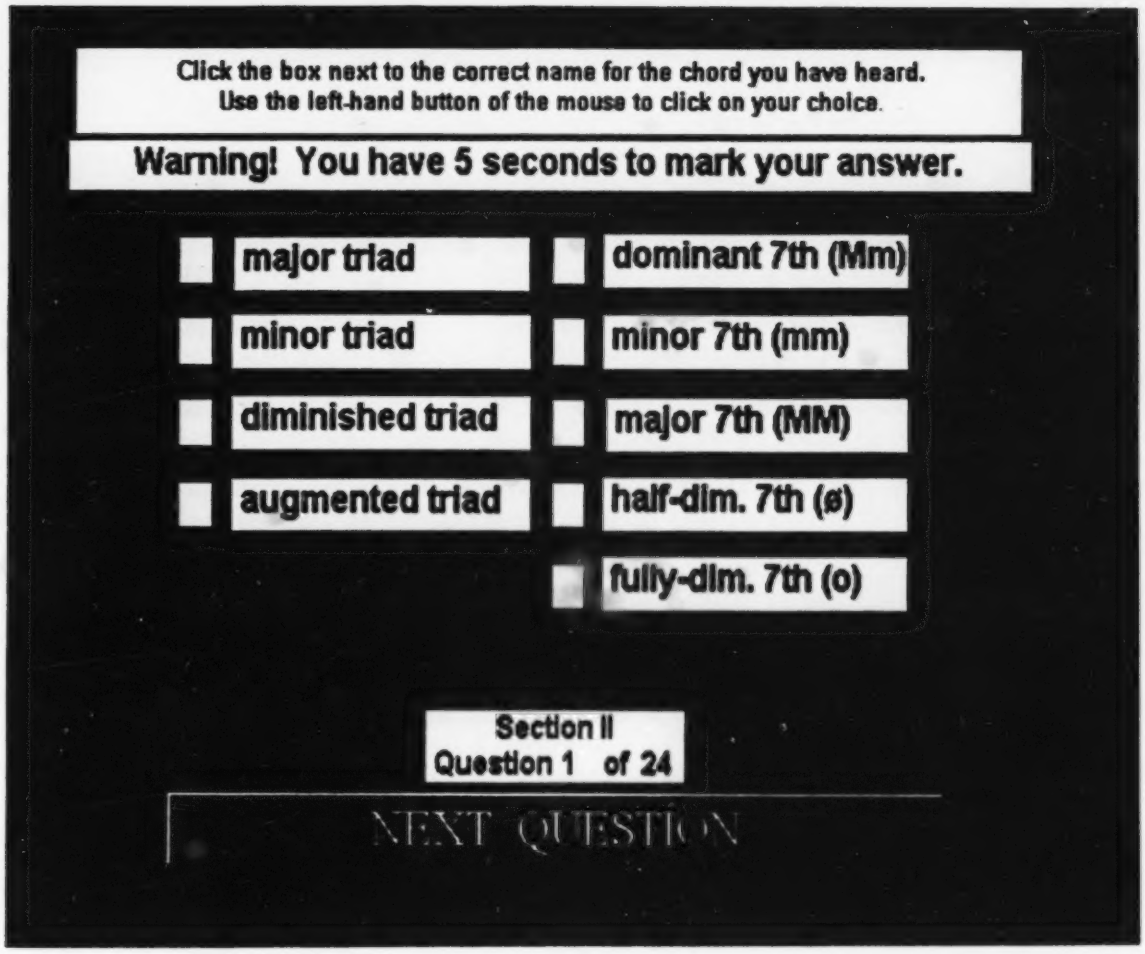


This screen for the sample question is identical to the screen presented for the Section II questions. Candidates hear an chord played on a piano over their headphones and then click on the correct identification of the chord they have heard. There is one sample question.



The message at the top of the screen appears when the test taker completes the sample question. The test taker is given the option of repeating the sample or going on to Section II.


Click the box next to the correct name for the chord you have heard.
Use the left-hand button of the mouse to click on your choice.
major triad
minor triad
diminished triad
augmented triad
dominant 7th (Mm)
minor 7th (mm)
major 7th (MM)
half-dim. 7th (ø)
fully-dim. 7th (o)
Section II
Question 1 of 24
NEXT QUESTION



The warning in the box at the top of the screen appears for each question in Sections I and II when sufficient time has elapsed since the playing of the stimulus for the question.

Directions for Section III – Harmonic Dictation

In this section you will hear two harmonic progressions. Each consists of 7 chords, voiced SATB. **Each progression in this section requires two actions**: you must notate the soprano and bass lines of the progression **AND** you must complete a Roman numeral analysis of each progression.

PART 1: You will have 2 minutes to complete each notation. The time remaining appears at the bottom left of the display. On the screen you will see a piano staff with the key signature, the meter signature and the opening soprano and bass notes given. You are to notate the soprano and bass lines ONLY of the progression you hear.

You may hear the progression up to 3 times. After the first playing, you will see a bar labeled **"Play Dictation"**. Click on this button to hear the dictation a second or third time.

PART 2: After 2 minutes have elapsed, you will see a set of instructions for completing the Roman numeral analysis, and a bar labeled **"Begin Analysis"**. When you click this bar, your notation will reappear and you will hear the progression once more. You will have 70 seconds to complete the analysis. During this time you may make additional changes to your notation. When time expires, your notation and analysis will automatically be recorded as they appear on the screen.

When you have finished reading these directions, click on the bar labeled **"View More Directions"** below to view directions on entering pitches.

View More Directions

Directions for Section III – Harmonic Dictation
- continued -

To notate a pitch on either staff, position the arrow on a line or space of the staff and click. USE THE LEFT-HAND BUTTON ON THE MOUSE TO POSITION THE NOTES. In each measure, you may enter only the number of notes appropriate for the meter signature.

To change a pitch on either staff, position the arrow over the note you wish to change. The note will automatically be highlighted. Now position the arrow, within the highlighted area, to the line or space you choose and click. The note will be changed to the pitch you indicated.

To delete a pitch, position the arrow over the note you wish to delete and press the delete key.

To insert accidental signs, first position the arrow to highlight the note you wish to alter. Type **S** to insert a sharp. Type **F** to insert a flat. Type **N** to insert a natural. Follow the same procedure to remove accidentals.

When you have finished reading these directions, click on the bar labeled **"Begin Practice Session"** below so that you may practice placing notes on the staff.

View Previous Screen | Begin Practice Session

Practice Screen for Section III
Harmonic Dictation

Practice entering and altering pitches on the staff below. When you are comfortable, click the bar labeled **"Begin Sample Question"** to try a sample Harmonic Dictation. Please note that the time allotted to the sample question is less than that allotted to the actual test questions.

Add a pitch – position the arrow on the staff and click the left-hand mouse button.

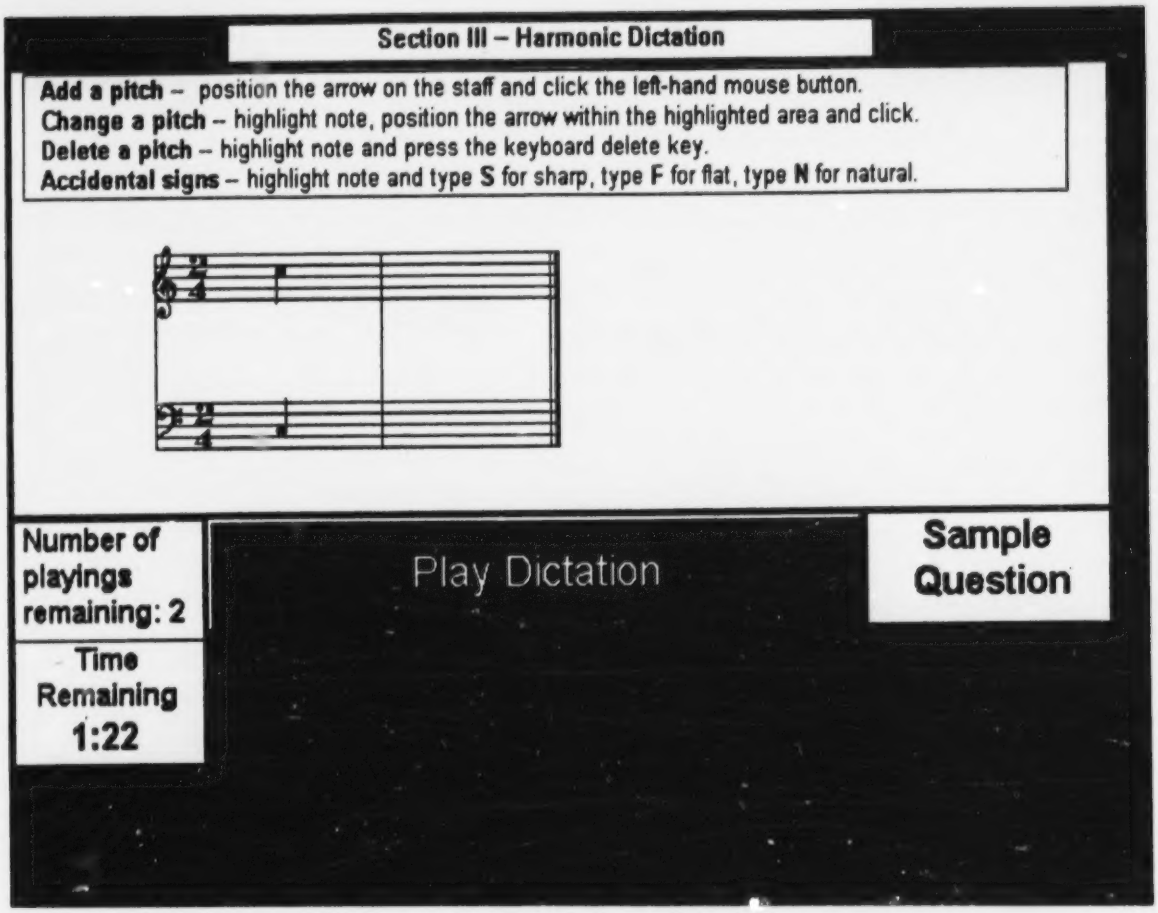
Change a pitch – highlight note, position the arrow within the highlighted area and click.

Delete a pitch – highlight note and press the keyboard delete key.

Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Begin Sample Question

This musical staff and the directions for inserting and changing notes is identical to that used for the questions in Section III. Test takers were given unlimited time to orient themselves to the software at this point.



Test takers hear an abbreviated version of the items in Section III--four chords instead of seven. They are given the opportunity to follow the same directions as those for the actual Section III exercises. The boxes on the lower left appear throughout the exercise in Sections III and IV. The abbreviated instructions for note entry appear at the top of each screen in these two sections.

Section III – Harmonic Dictation

Roman Numeral Analysis

After viewing these instructions your notation will reappear on the screen. In addition, you will see columns of Roman numeral analysis symbols beneath chords 2 through 5.

The progression will be played one final time. You are to click on the symbol in each column that correctly identifies the chord you hear. When you click on your choice, that symbol will be highlighted. You may change your choice by clicking on another symbol in the same vertical column. **CONSIDER THE CHOICES FOR EACH CHORD SEPARATELY**. The analysis for chords 1 and 7 is given.

You will have 30 seconds to complete this analysis. You may use this time to make changes to your notation, as well. When time expires, your notation and analysis will automatically be recorded as they appear on the screen.

When you have finished reading these directions, click on the bar labeled **"Begin Analysis"** below.

Number of playings remaining: 1

Time Remaining 0:00

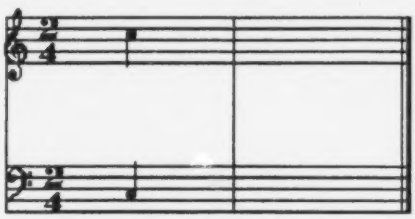
Begin Analysis

Sample Question

The second part of the Harmonic Dictation exercise is introduced and explained here, after test takers have become familiar with the note entry portion of the exercise. The same sample exercise appears on the opposite page.

Section III – Harmonic Dictation

Add a pitch -- position the arrow on the staff and click the left-hand mouse button.
Change a pitch -- highlight note, position the arrow within the highlighted area and click.
Delete a pitch -- highlight note and press the keyboard delete key.
Accidental signs -- highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.



Number of playings remaining: 0

Time Remaining 0:20

I

V

V

V/V

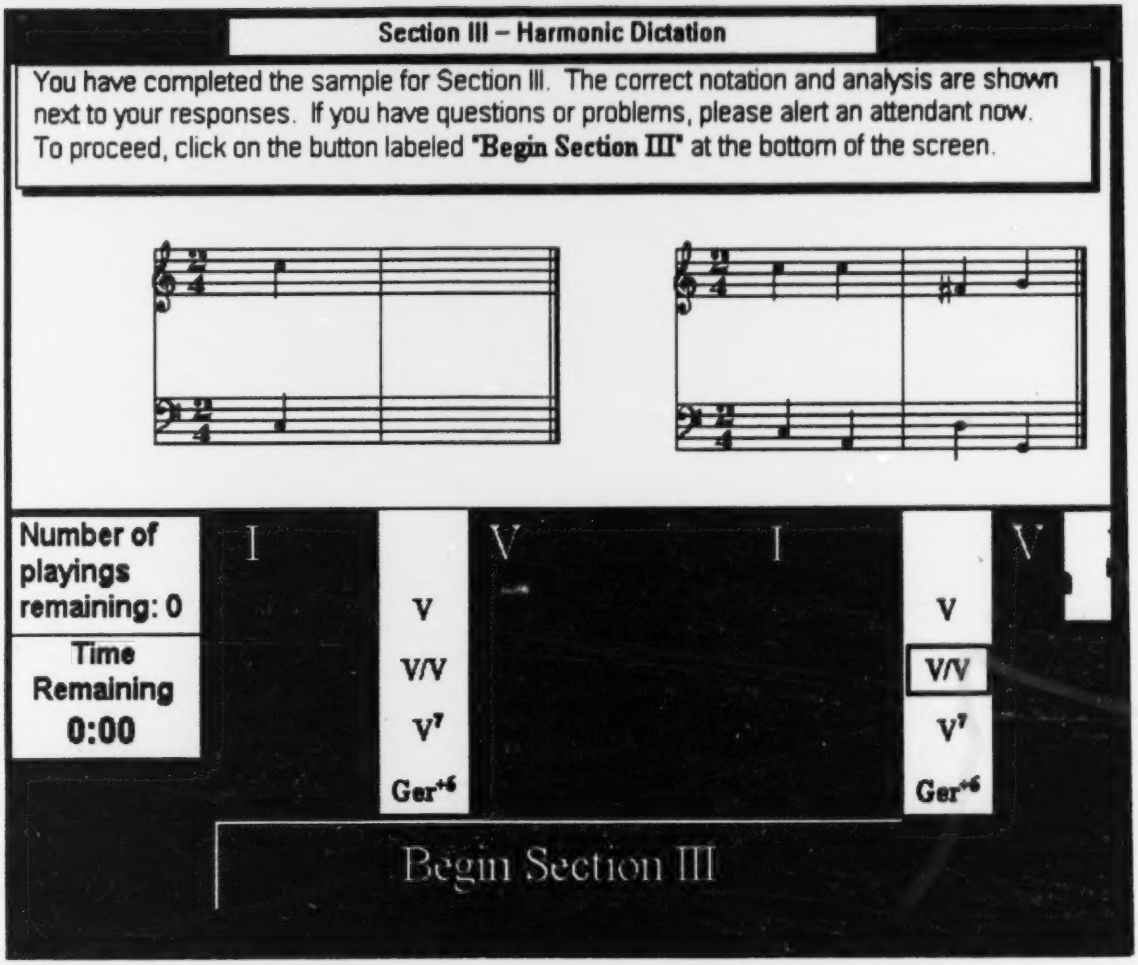
V^7

Ger^{+6}

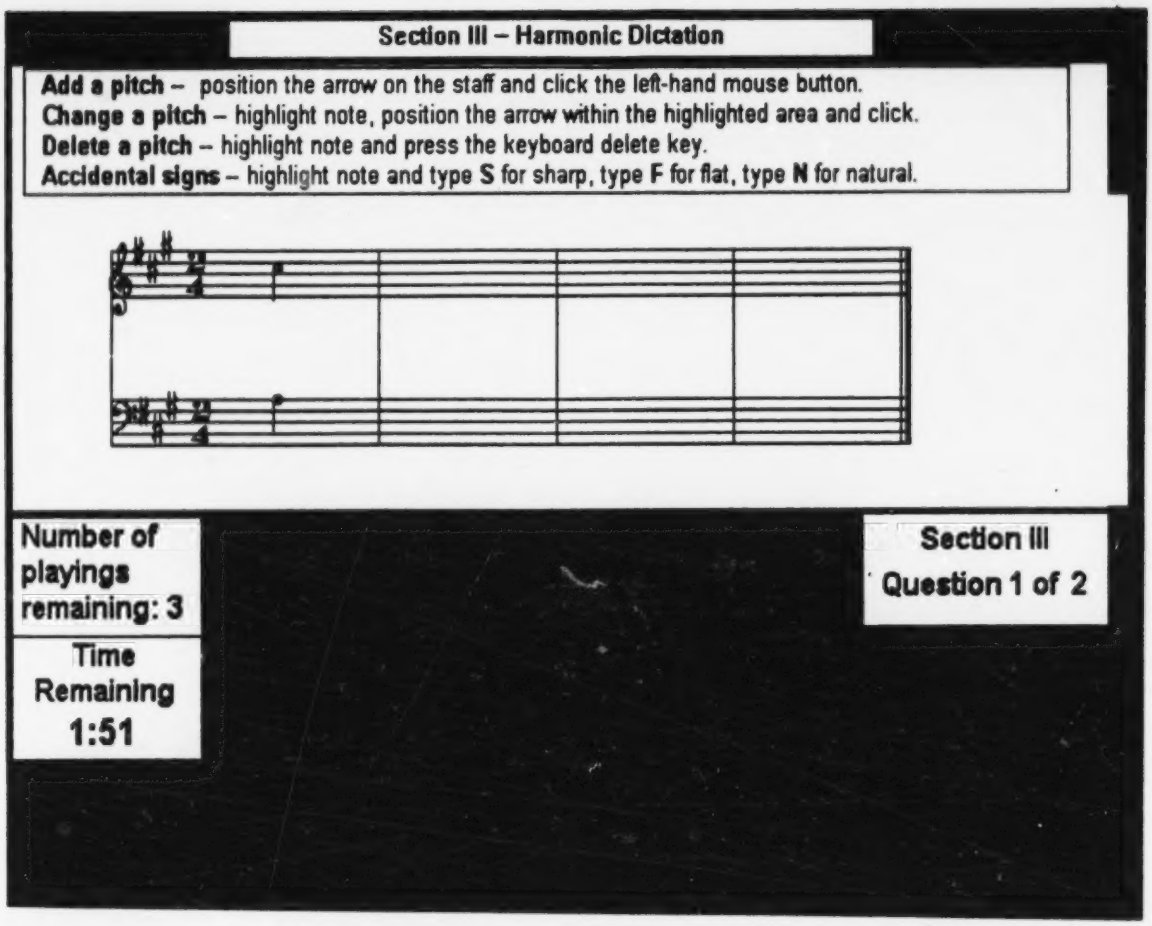
Sample Question

Click on **one** Roman numeral in each white or yellow column. A box will appear marking your choice. Consider the Roman numeral analysis for **each chord separately**.

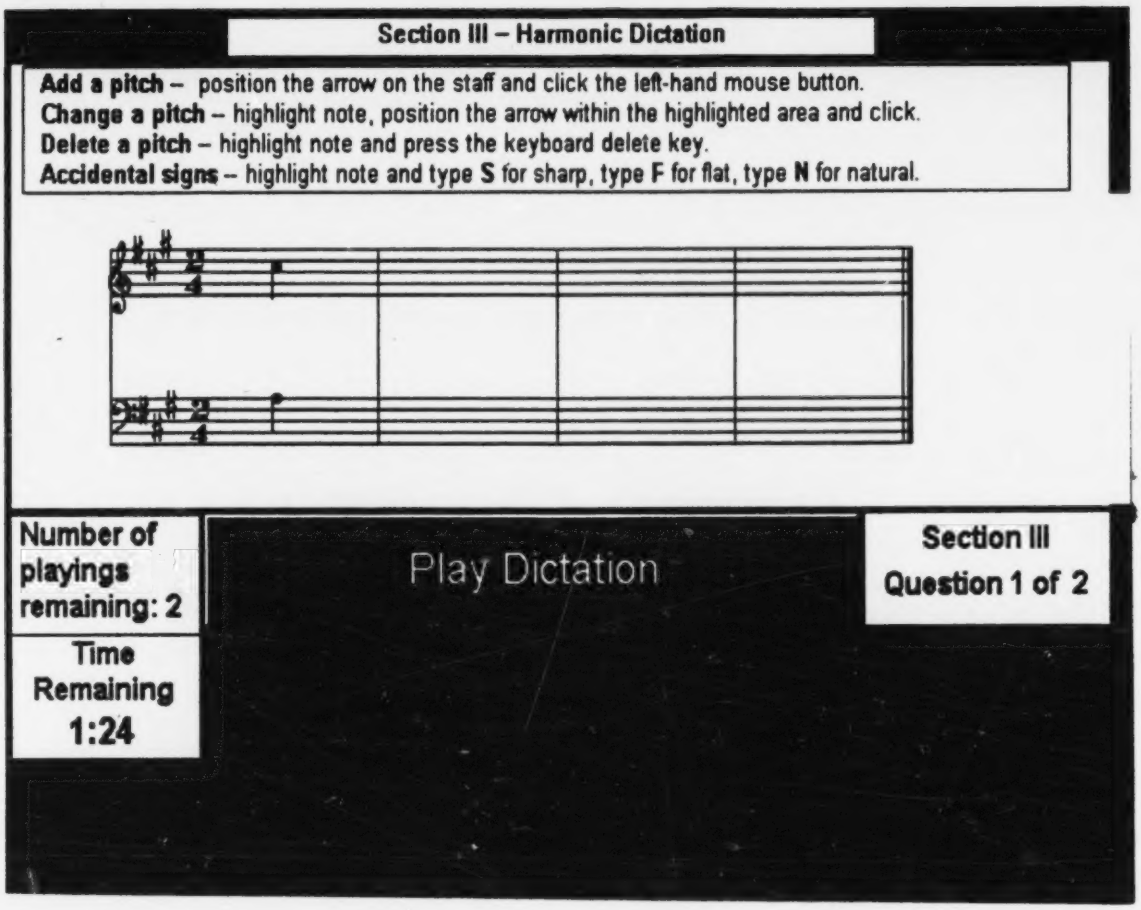
The box at the bottom of the screen summarizes the directions for this portion of the exercise in this section.



The sample question response is presented on the left; the correct response on the right. Test takers are given as long as they wish to study the two responses.



This screen appears after the test taker clicks on Begin Section III at the bottom of the previous screen. The aural stimulus, the dictation, is played automatically for the first time. The boxes at the lower left record the number of playings remaining and the total time remaining. Test takers choose the pace at which they hear the aural stimulus repeated within the total time allotted for the exercise.



After the initial playing the message at the center of the screen appears, allowing test takers to repeat the stimulus at will.

Section III – Harmonic Dictation

Roman Numeral Analysis

After viewing these instructions your notation will reappear on the screen. In addition, you will see columns of Roman numeral analysis symbols beneath chords 2 through 5.

The progression will be played one final time. You are to click on the symbol in each column that correctly identifies the chord you hear. When you click on your choice, that symbol will be highlighted. You may change your choice by clicking on another symbol in the same vertical column. **CONSIDER THE CHOICES FOR EACH CHORD SEPARATELY**. The analysis for chords 1 and 7 is given.

You will have 70 seconds to complete this analysis. You may use this time to make changes to your notation, as well. When time expires, your notation and analysis will automatically be recorded as they appear on the screen.

When you have finished reading these directions, click on the bar labeled **"Begin Analysis"** below.

Number of playings remaining: 1

Time Remaining 0:00

Begin Analysis

Section III
Question 1 of 2

This screen gives the directions for the second part of the harmonic dictation exercise. Test takers are given unlimited time to read these instructions.

Section III – Harmonic Dictation

Add a pitch – position the arrow on the staff and click the left-hand mouse button.
Change a pitch – highlight note, position the arrow within the highlighted area and click.
Delete a pitch – highlight note and press the keyboard delete key.
Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Number of playings remaining: 1

Time Remaining 1:08

IV^6	It^{+6}
ii	V
V	V^7/V
vi	I^6_4

Section III
Question 1 of 2

Click on **one** Roman numeral in each white or yellow column. A box will appear marking your choice. Consider the Roman numeral analysis for **each chord separately**.

The test taker's notation of the dictation appears on the screen above the analysis boxes. The aural stimulus is played one more time. The test taker can change the notation as well as choose the Roman Numeral analysis beneath the designated chords. The clock is visible on the screen at all times.

Section III – Harmonic Dictation

Add a pitch -- position the arrow on the staff and click the left-hand mouse button.
Change a pitch -- highlight note, position the arrow within the highlighted area and click.
Delete a pitch -- highlight note and press the keyboard delete key.
Accidental signs -- highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Number of playings remaining: 1	I			I	Section III Question 2 of 2
Time Remaining 1:05		vi	V^7		
		V	V		
		IV^6	ii^6		
		I^6	I^6_4		

Click on **one** Roman numeral in each white or yellow column. A box will appear marking your choice. Consider the Roman numeral analysis for **each chord separately**.

This is the shell for the second of the harmonic dictation exercises.

Directions for Section IV – Melodic Dictation

In this section you will hear two melodic dictations. On the screen you will see a treble or bass clef staff with the key signature, the meter signature and the first note of the dictation given. You are to notate the dictation you hear.

After the first playing, you will see a bar labeled **"Play Dictation"**. Click on this bar to hear the dictation again. You may hear the dictation up to 4 times. This bar will not be visible on the screen while the dictation is playing. You will have a total of 3 minutes and 30 seconds to complete your notation. The time remaining appears at the bottom left of the display.

To hear your own notation, click the bar labeled **"Play Your Response"**. This bar will be visible after the first playing of the dictation. Note that **ONCE YOU PLAY BACK YOUR RESPONSE, YOU MAY NOT HEAR THE DICTATION AGAIN.**

When you have finished reading these directions, click on the bar labeled **"View More Directions"** to view directions on notating pitches on the staff. Directions for placing notes on the staff, changing pitches, and inserting accidentals are identical to those for the previous section (Harmonic Dictation).

New techniques for Melodic Dictation include using the mouse to change note values and to delete pitches from the staff.

View More Directions

On this screen and the one reproduced on the facing page, the directions for melodic dictation are presented. All of the functions used in harmonic dictation remain unchanged, and some others are added.

Directions for Section IV – Melodic Dictation
- continued -

To notate a pitch on the staff, position the arrow on a line or space of the staff and click. USE THE LEFT-HAND BUTTON ON THE MOUSE TO POSITION THE NOTES.

To change a pitch on the staff, position the arrow over the note you wish to change. The note will automatically be highlighted. Now position the arrow, within the highlighted area, to the line or space you choose and click. The note will be changed to the pitch you indicated.

To delete a pitch, position the arrow over the note you wish to delete and press the delete key.

To insert accidental signs, first position the arrow to highlight the note you wish to alter. Type **S** to insert a sharp. Type **F** to insert a flat. Type **N** to insert a natural. Follow the same procedure to remove accidentals.

To change the value of a note, position the arrow over the note so that it is highlighted and click the right-hand button on the mouse. The note values follow this cycle: eighth, quarter, half, whole. Keep clicking until the highlighted note acquires the desired value.

When you have finished reading these directions, click on the bar labeled "Begin Sample Session" below so that you may practice placing notes on the staff for the melodic dictation section.

View Previous Screen | Begin Sample Session

BEST COPY AVAILABLE

Practice Screen for Section IV – Melodic Dictation

Practice entering and altering pitches on the staff below. When you are comfortable, click the button labeled "Begin Sample Question" to try a sample Harmonic Dictation. Please note that the time allotted to the sample question is less than that allotted to the actual test questions.

Add a pitch – position the arrow on the staff and click the left-hand mouse button.

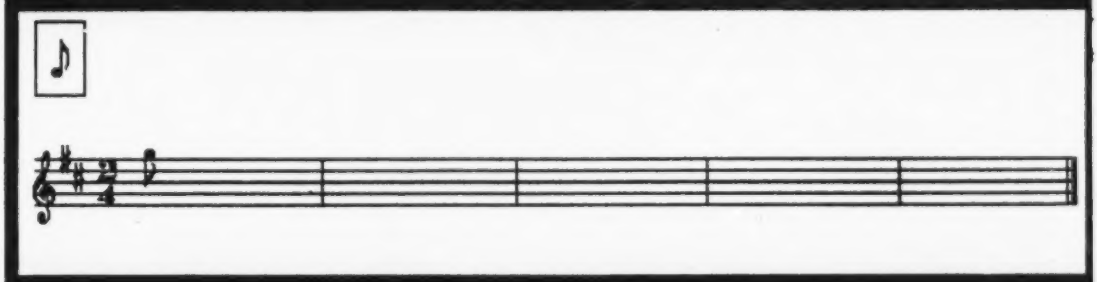
Change a pitch – highlight note, position the arrow within the highlighted area and click.

Delete a pitch – highlight note and press the keyboard delete key.

Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Change value of a note – highlight note and click RIGHT-HAND button on the mouse.

Note values cycle: eighth, quarter, half, whole. Keep clicking until note acquires desired value. The current note value is displayed in the box over the clef sign.



Begin Sample Question

A sample exercise is provided for the test taker, just as in the previous section. All of the features that characterize the actual exercise--stimulus repetitions, note entry, time clock, and playback--are available in the sample exercise.

Section IV – Melodic Dictation

Add a pitch – position the arrow on the staff and click the left-hand mouse button.

Change a pitch – highlight note, position the arrow within the highlighted area and click.

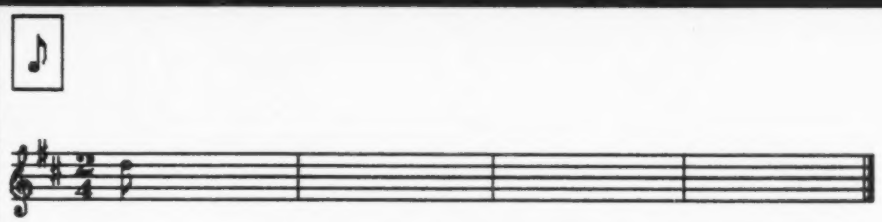
Delete a pitch – highlight note and press the keyboard delete key.

Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Change value of a note – highlight note and click RIGHT-HAND button on the mouse.

Note values cycle: eighth, quarter, half, whole. Keep clicking until note acquires desired value.

The current note value is displayed in the box over the clef sign.



Number of playings remaining: 3	Play Dictation	**Sample Question**
Time Remaining 2:07	Play Your Response	

In the melodic dictation section, test takers are allowed to hear their own notation of the stimulus. They may choose to hear it at any time after notes are entered, but once the test taker's response is played back, the stimulus cannot be repeated. The box at the bottom of the screen controls the playback of the test taker's response.

Section IV – Melodic Dictation

Add a pitch – position the arrow on the staff and click the left-hand mouse button.

Change a pitch – highlight note, position the arrow within the highlighted area and click.

Delete a pitch – highlight note and press the keyboard delete key.

Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Change value of a note – highlight note and click RIGHT-HAND button on the mouse.

Note values cycle: eighth, quarter, half, whole. Keep clicking until note acquires desired value.

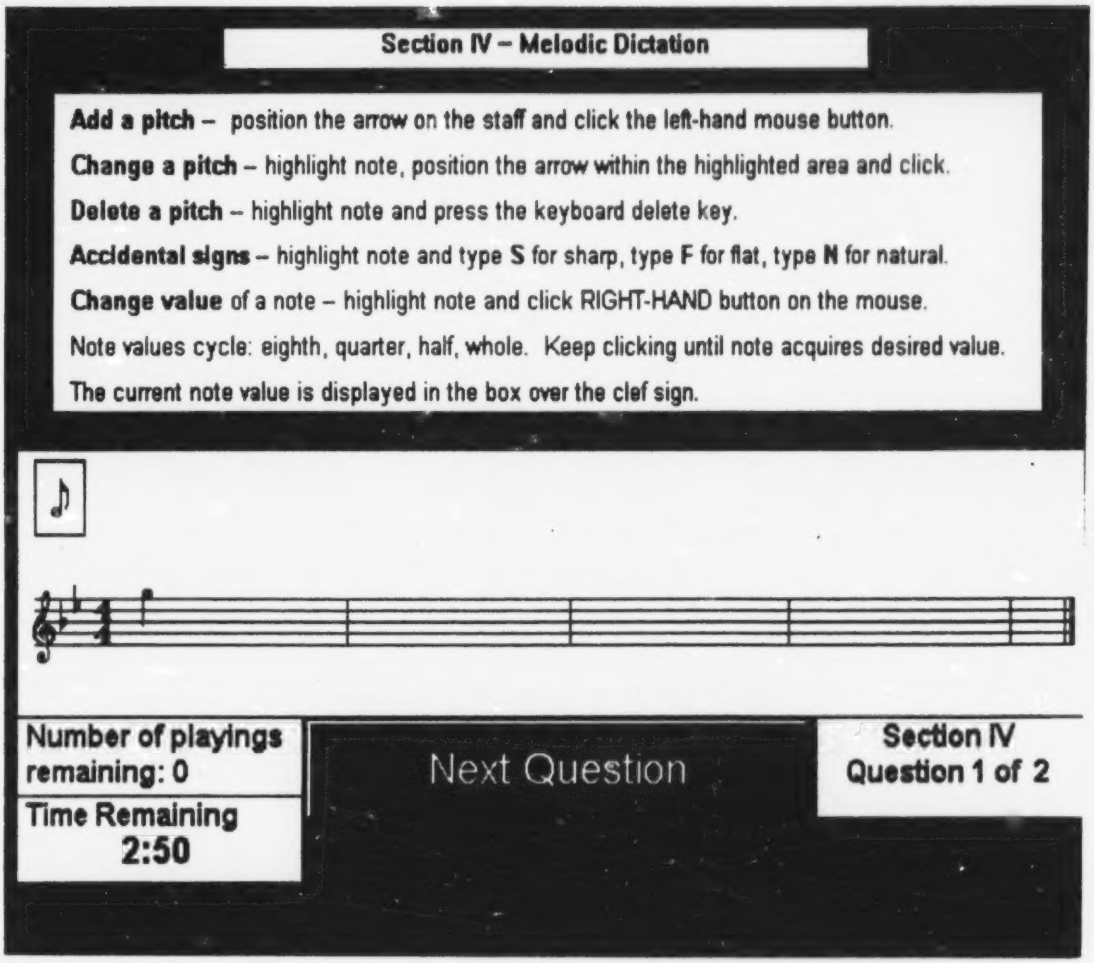
The current note value is displayed in the box over the clef sign.

Number of playings remaining: 3

Time Remaining
3:21

Section IV
Question 1 of 2

The melodic dictation exercise screen, like the harmonic dictation exercise screen, displays the number of playings available and the countdown clock. The box to the left of the screen, above the clef sign on the musical staff, shows the value of the note currently selected. The second mouse button is used to change the value of a note. Melodic dictation is the only exercise in which note values (and thus rhythm) were part of the exercise.



In the center of the screen the Next Question command allows the test taker to advance at will to the second exercise in this section.

Section IV – Melodic Dictation

Add a pitch – position the arrow on the staff and click the left-hand mouse button.

Change a pitch – highlight note, position the arrow within the highlighted area and click.

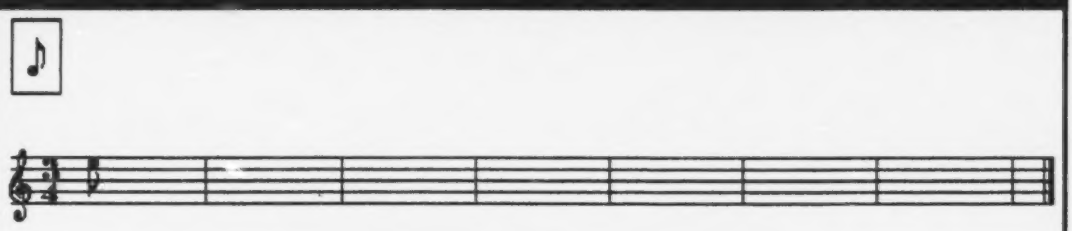
Delete a pitch – highlight note and press the keyboard delete key.

Accidental signs – highlight note and type **S** for sharp, type **F** for flat, type **N** for natural.

Change value of a note – highlight note and click RIGHT-HAND button on the mouse.

Note values cycle: eighth, quarter, half, whole. Keep clicking until note acquires desired value.

The current note value is displayed in the box over the clef sign.



Number of playings remaining: 3

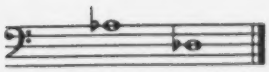
Time Remaining 3:25

Section IV Question 2 of 2

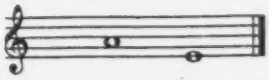
This screen displays the second of the melodic dictation exercises.

Section 1: Intervals
(Aural stimuli)

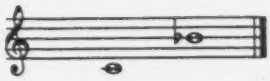
1



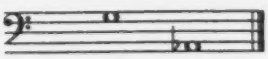
2



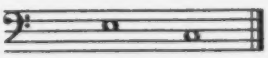
3



4



5



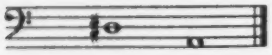
6

7

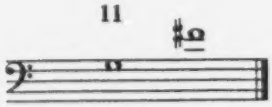
8

9

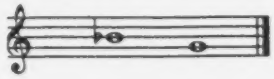
10



11



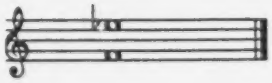
12



13

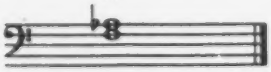
14

15

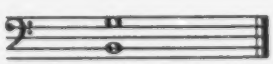


16

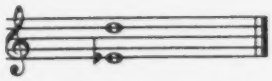
17



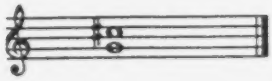
18



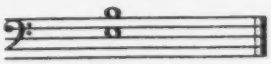
19



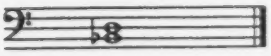
20



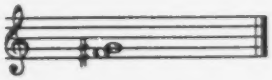
21



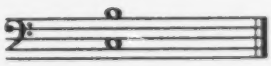
22



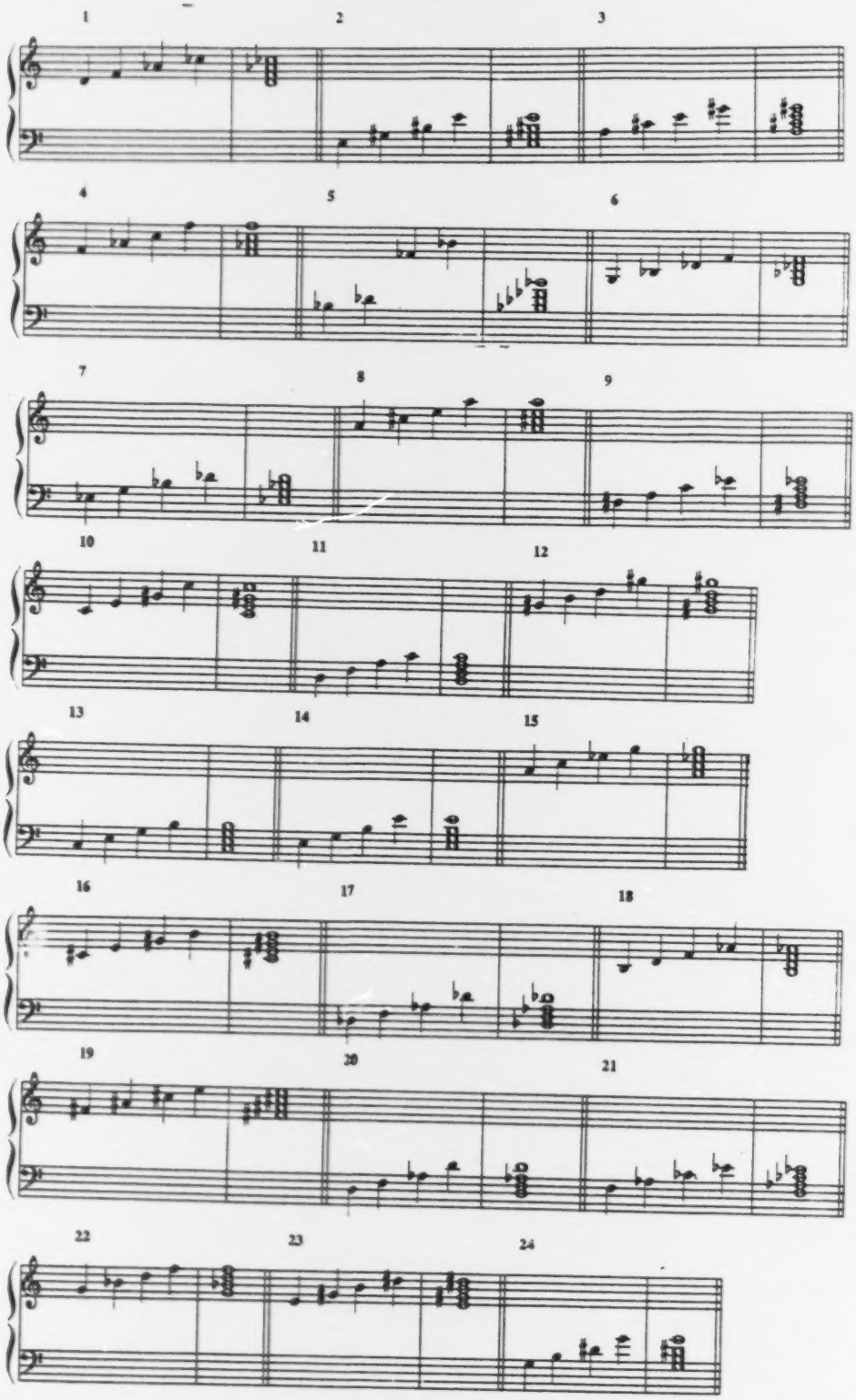
23



24

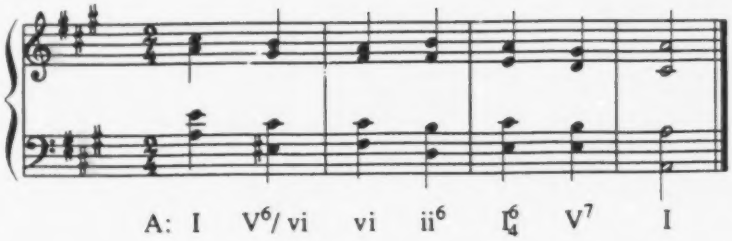


Section 2: Chords
(Aural Stimuli)

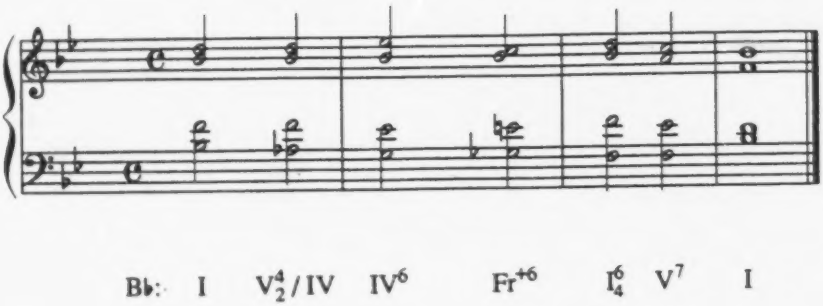


Section 3: Harmonic Dictation

Harmonic Dictation 1



Harmonic Dictation 2



Section 4: Melodic Dictation

Melodic Dictation 1

Melodic Dictation 2

APPENDIX B

The questionnaire filled out by each test taker after the test administration is reproduced below. The salient features of test takers' responses are incorporated into the report text.

Feedback Questionnaire for PRPC pilot testing

June 10-12, 1992

Name

Address (through July)

Social Security #

Year in school

Major

What aural skills courses have you completed?

What music theory courses have you completed?

Have you ever used a computer before?

Have you ever used a mouse?

Have you ever used a computer to respond to or manipulate musical sounds or symbols?

For each question or set of questions below, please indicate your response by checking one of the given choices. If you wish to amplify your response, use the space below the choices.

1. **Timing**

Intervals	Adequate □	Inadequate □
Chords	Adequate □	Inadequate □
Harmonic Dictation	Adequate □	Inadequate □
Melodic Dictation	Adequate □	Inadequate □

For which of the four sections, if any, did you feel that you were given more time than you needed to answer?

Intervals □

Chords □

Harmonic Dictation □

Melodic Dictation □

2. **Mechanics**

Were you comfortable working with the mouse in each of these sections?
Please comment on all NO answers.

Intervals	Yes □	No □

Comment:

Chords	Yes □	No □

Comment:

Harmonic Dictation Yes □ No □

Comment:

Melodic Dictation Yes □ No □

Comment:

Evaluate the ease of the following actions you were expected to perform with the mouse in the harmonic and melodic dictation exercises:

clicking the notes into the proper measure
Easy □ Hard □

clicking on the space or line where you wanted the note to appear
Easy □ Hard □

changing the pitch of a note when necessary
Easy □ Hard □

clicking on the note to obtain the correct note value
Easy □ Hard □

3. **Feedback**

Was the feedback from the computer as you took the sample questions adequate and helpful? Please comment on any NO answer.

Intervals Yes □ No □

Comment:

Chords Yes □ No □

Comment:

Harmonic Dictation Yes □ No □

Comment:

Melodic Dictation Yes □ No □

Comment:

Would you have liked more feedback as you took the sample questions?
Please explain all YES answers.

Intervals Yes □ No □

Comment:

Chords Yes □ No □

Comment:

Harmonic Dictation Yes □ No □

Comment:

Melodic Dictation Yes □ No □

Comment:

Below, briefly comment on the playback feature of the melodic dictation exercises.

4. **Test difficulty**

Considering the questions themselves (apart from the computer format), in each section, did you find them easy, medium, hard, or very difficult?

Intervals Easy □ Medium □ Hard □ Very difficult □

Comment:

Chords Easy □ Medium □ Hard □ Very difficult □

Comment:

Harmonic Dictation Easy □ Medium □ Hard □ Very difficult □

Comment:

Melodic Dictation Easy □ Medium □ Hard □ Very difficult □

Comment:

5. Compare this test in each of the following ways to the same test given using paper and pencil:

This test was [harder, easier] to take.

The directions for this test were [clearer, more confusing].

The computer format [did, did not] affect the level of my performance.

This test has [advantages, disadvantages] in comparison with a traditional paper and pencil test. Specify what they are.

APPENDIX C

Below are a number of graphic presentations of the performance data on the experimental test. We have used only the second administration of the test, since we believe that the first administration data is not particularly useful. In all these data, the five test takers who experienced hardware failure during the administration are excluded because their scores are necessarily inaccurate.

Table 1 presents the scores of the 65 test takers in descending order by total score. The breakdown of each total score by individual section is given for each test taker. Tables 2 and 3 present frequency distributions by total score and by section scores, respectively. Figures 1-10 present the frequency distributions of scores by total and by each section, and distributions by test taker. A comparison of these graphs will reveal the asymmetrical distribution of scores in the Intervals and Chords sections as compared to the Harmonic and Melodic Dictation sections. It is true that the aural skills tested in the Intervals and Chords sections are less sophisticated than those tested in the two Dictation sections, but they are very closely related. Most musicians would agree that the scores would likely be much more similar than was the case here. This suggests the role played by unfamiliarity with the software and the computer itself in a testing situation. It is notable that the distribution by individual for Melodic Dictation (Figure 10) shows the most precipitous decline in performance very early in the sample. This was the most difficult section for this group of test takers.

Table 1: Examinee Scores and Section Scores by Total Score

EXAMINEE	Total Score (max. 139)	Interval (max. 24)	Chords (max. 24)	Harm.Dict. (max. 35)	Mel. Dict. (max. 56)
1	129	23	22	33	51
2	119	20	20	33	46
3	114	22	14	30	48
4	112	18	16	27	51
5	112	18	20	35	39
6	108	23	21	29	35
7	101	23	23	30	25
8	92	8	13	34	37
9	92	17	13	30	32
10	91	16	16	34	25
11	91	19	19	28	25
12	88	21	23	27	17
13	85	15	16	24	30
14	83	21	14	28	20
15	82	12	14	26	30
16	80	15	12	33	20
17	79	20	12	24	23
18	76	22	20	4	30
19	73	14	10	22	27
20	73	19	15	24	15
21	72	17	13	24	18
22	72	15	13	27	17
23	71	19	8	21	23
24	70	20	14	27	9
25	69	13	17	25	14
26	69	11	17	23	18
27	69	18	21	13	17
28	69	15	12	21	21
29	68	18	14	18	18
30	67	19	16	27	5
31	66	10	8	29	19
32	65	14	5	25	21
33	65	18	10	19	18
34	64	17	11	16	20
35	64	20	18	17	9
36	64	12	12	24	16
37	62	13	14	19	16
38	61	17	15	13	16
39	60	11	9	28	12
40	59	12	7	22	18
41	58	15	11	18	14
42	58	14	13	18	13
43	58	9	8	18	23
44	56	15	9	23	9
45	56	18	7	13	18
46	54	13	9	20	12
47	51	10	8	24	9
48	50	7	3	20	20
49	49	12	5	14	18
50	49	11	6	16	16
51	47	5	9	24	9
52	47	13	9	16	9
53	47	8	11	7	21
54	45	11	3	24	7
55	45	5	3	21	16
56	43	7	4	21	11
57	42	13	11	11	7
58	39	9	8	9	13
59	39	6	7	16	10
60	34	7	10	8	9
61	34	8	6	12	8
62	33	8	3	10	12
63	32	8	6	16	2
64	32	5	3	16	8
65	24	5	1	12	6

Table 2: Frequency Distribution of Total Scores

Score Attained	Number of Examinees
129	1
119	1
114	1
112	2
108	1
101	1
92	2
91	2
88	1
85	1
83	1
82	1
80	1
79	1
76	1
73	2
72	2
71	1
70	1
69	4
68	1
67	1
66	1
65	2
64	3
62	1
61	1
60	1
59	1
58	3
56	2
54	1
51	1
50	1
49	2
47	3
45	2
43	1
42	1
39	2
34	2
33	1
32	2
24	1
Total	**65**

Table 3: Frequency Distribution of Scores by Section

Score Attained	Interval (max. 24)	Chords (max. 24)	Harmonic Dictation (max. 36)	Melodic Dictation (max. 56)
51				2
48				1
46				1
39				1
37				1
35			1	1
34			2	0
33			3	0
32			0	1
30			3	3
29			2	0
28			3	0
27			5	1
26			1	0
25			2	3
24	0	0	8	0
23	3	2	2	3
22	2	1	2	0
21	2	2	4	3
20	4	3	2	4
19	4	1	2	1
18	6	1	4	7
17	4	2	1	3
16	1	4	6	5
15	6	2	0	1
14	3	6	1	2
13	5	5	3	2
12	4	4	2	3
11	4	4	1	1
10	2	3	1	1
9	2	5	1	7
8	5	5	1	2
7	3	3	1	2
6	1	3	0	1
5	4	2	0	1
4	0	1	1	0
3	0	5	0	0
2	0	0	0	1
1	0	1	0	0
Total	**65**	**65**	**65**	**65**

Table 1:

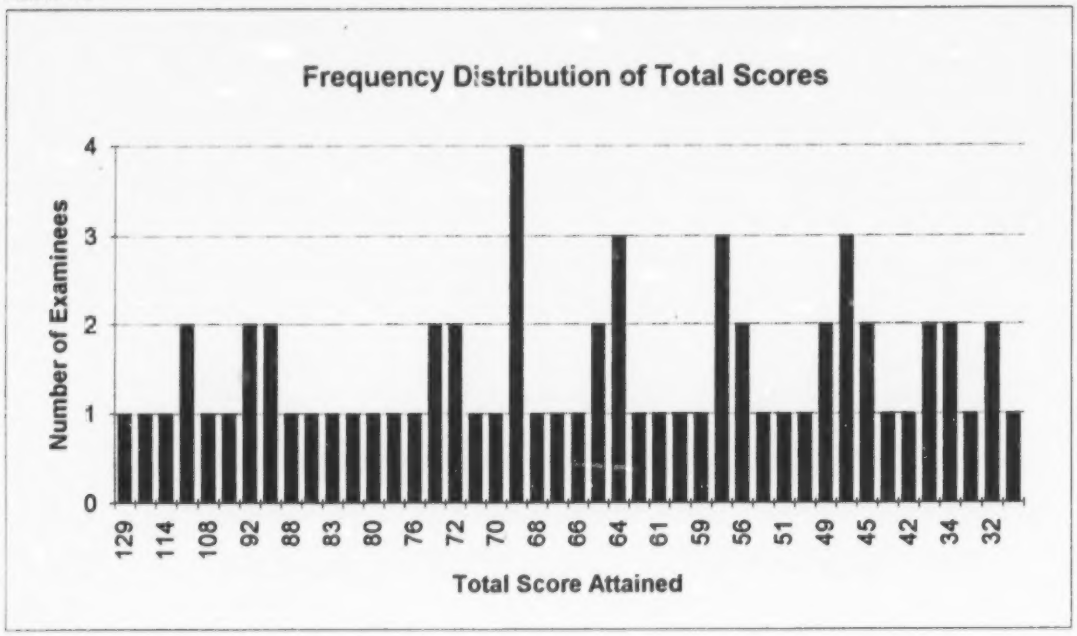


Table 2:

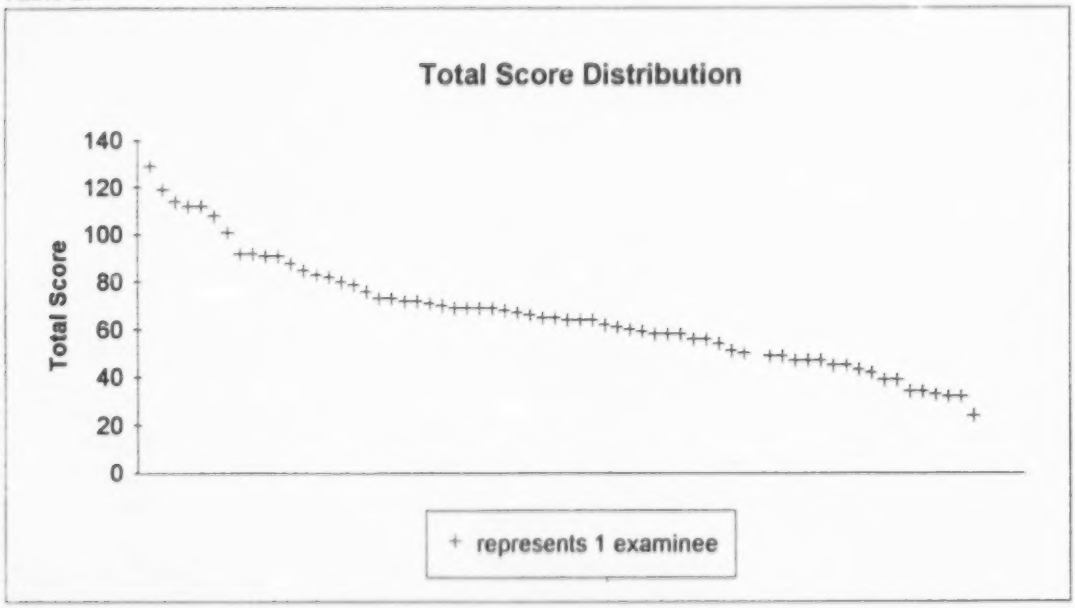


Table 3:

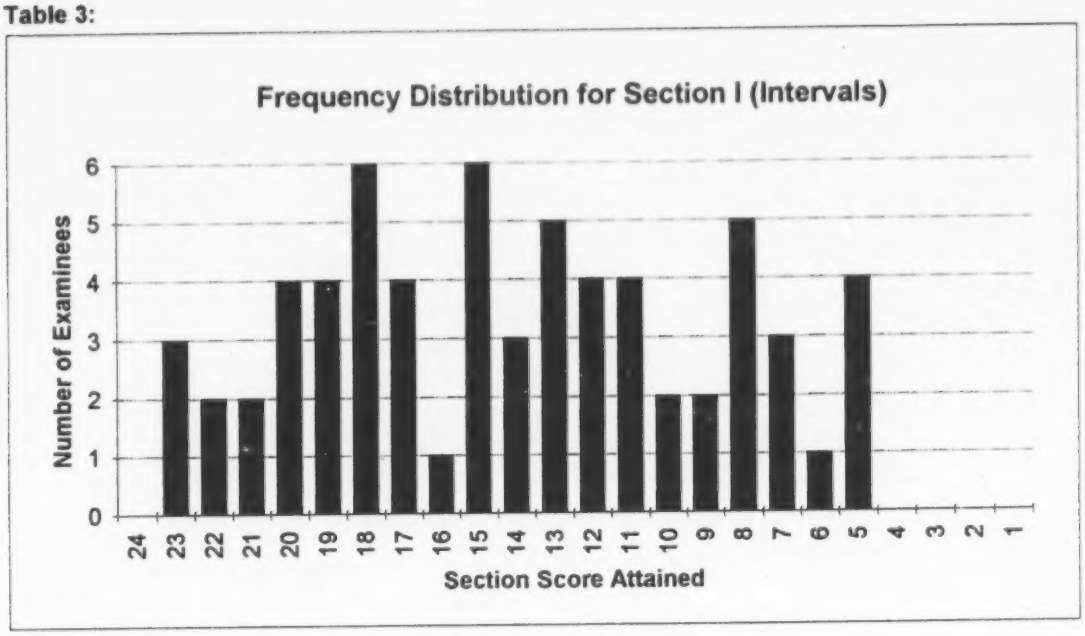


Table 4:

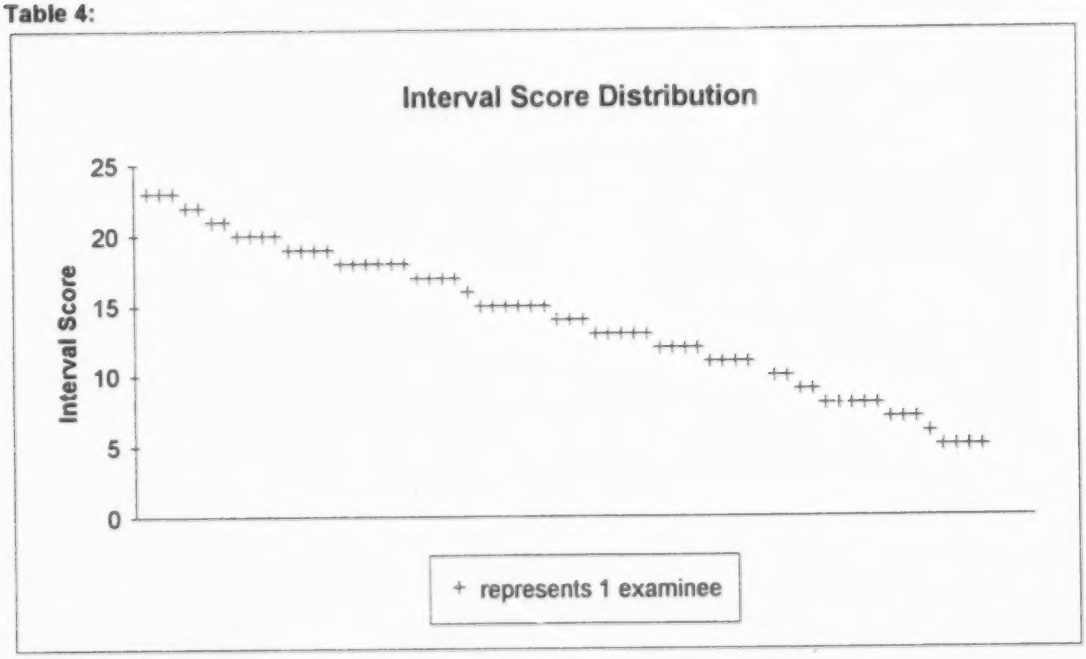


Table 5:

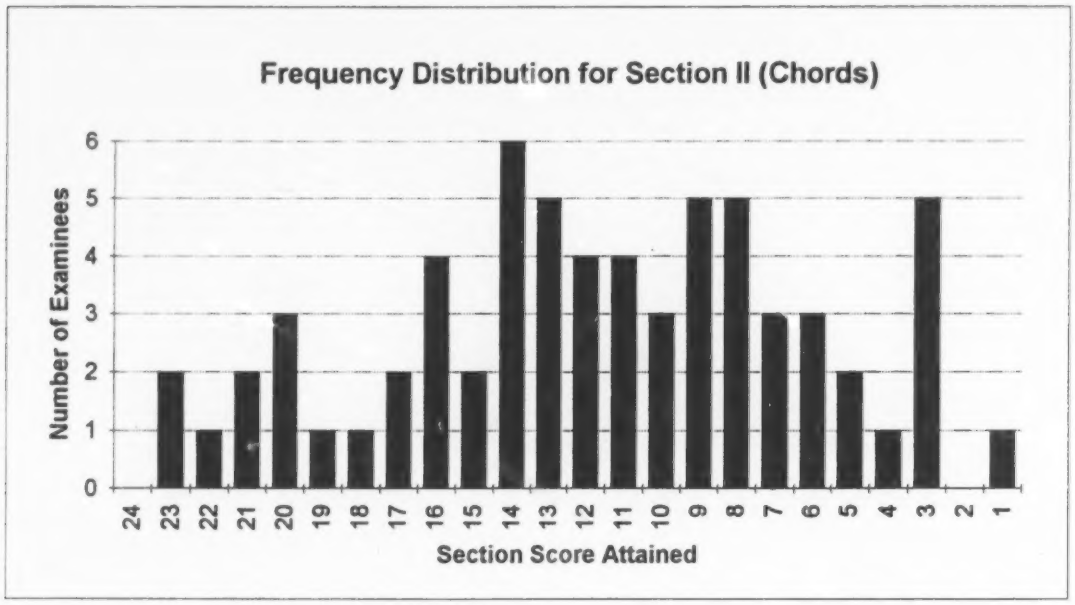


Table 6:

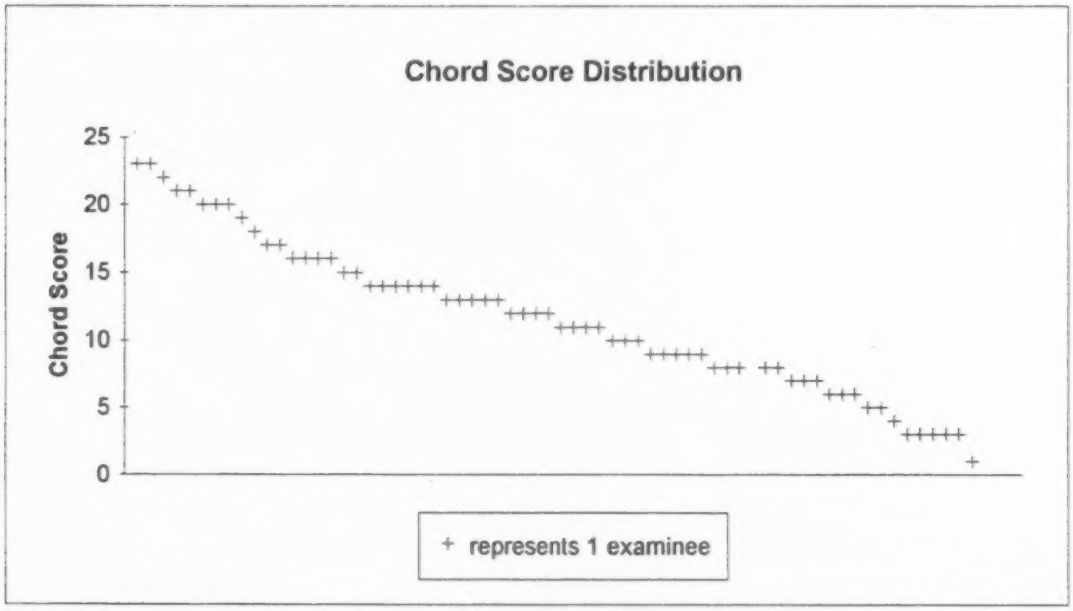


Table 7:

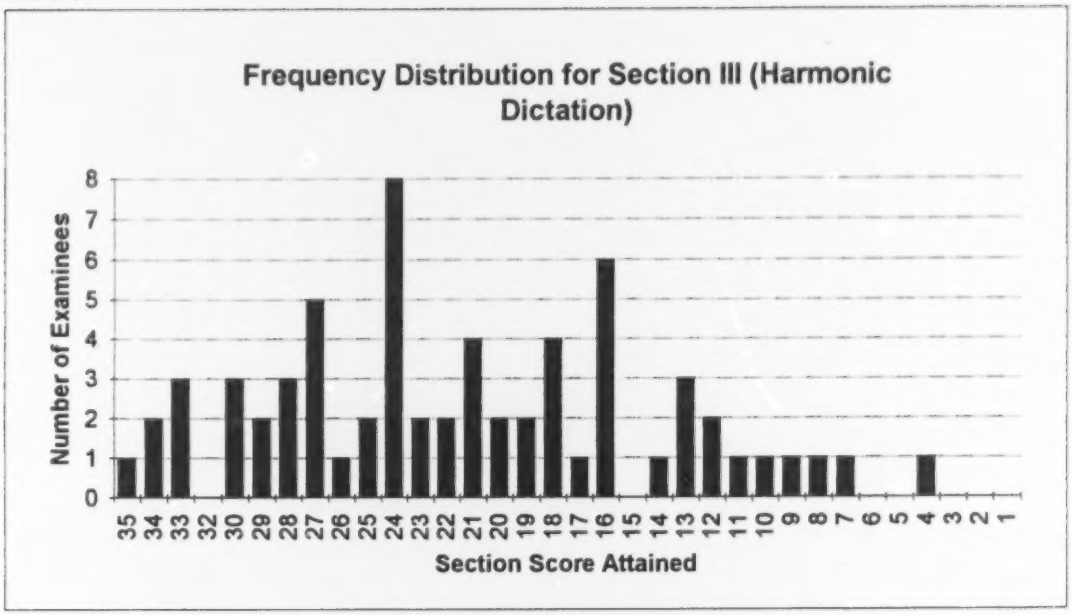


Table 8:

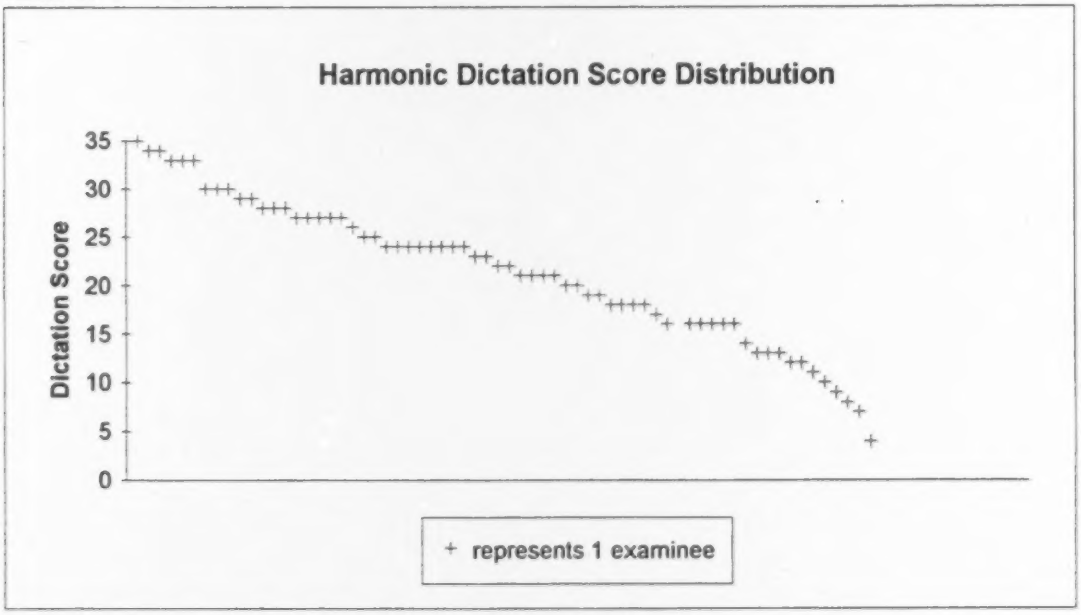


Table 9:

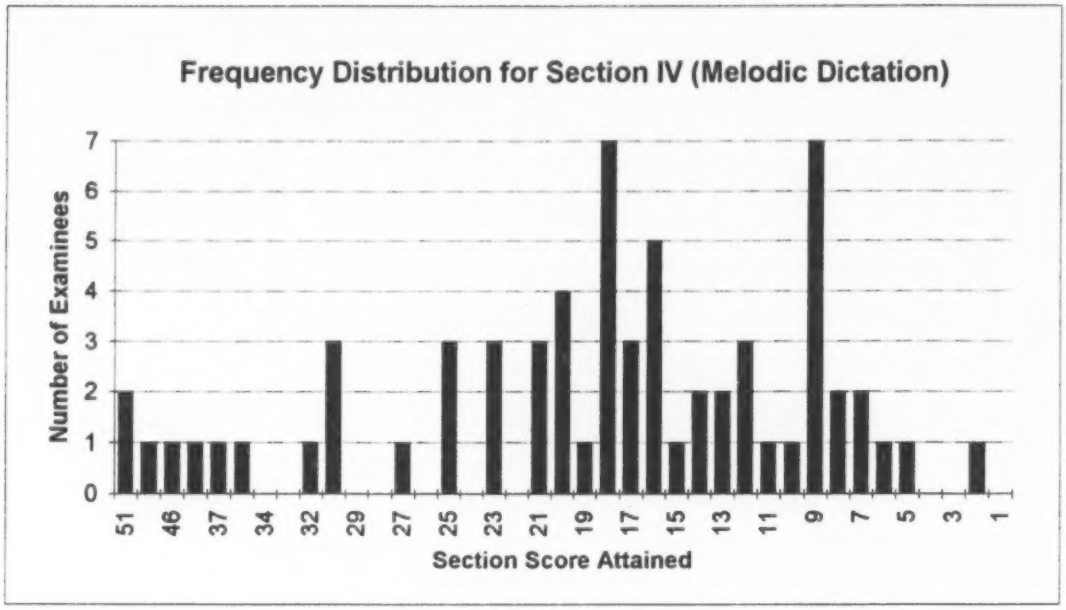
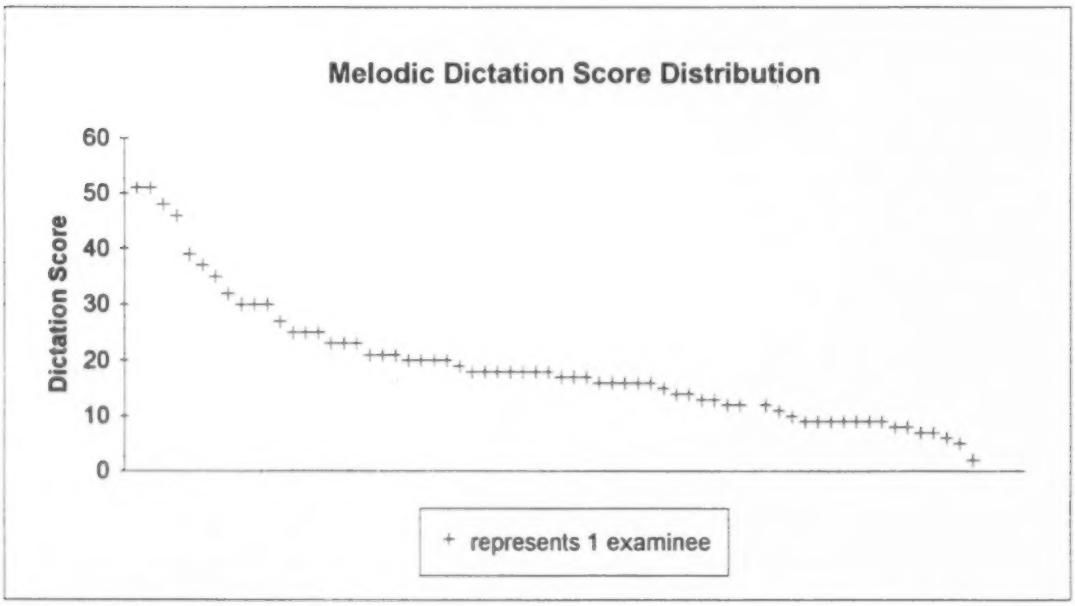


Table 10:



APPENDIX D

To supply some context for the scores on the experimental test and to at least attempt some measure of concurrent validity, we interviewed four professors at Westminster Choir College who had instructed many of the test takers in music theory and aural skills courses. We asked them for an analysis of the following for each test taker:

1. Achieved grades in written theory
2. Achieved grades in aural skills (sight singing and ear training)
3. An anecdotal analysis of the test taker's abilities (as contrasted with achievements) and any special information that would be of assistance.

For most test takers, this information was simply a confirmation of the test results, as Table 4 below shows, but in some cases the test results were surprising in light of either the test taker's achievements or abilities as perceived by his/her teachers. It is interesting (and reassuring) to note that the test results seem to confirm anecdotal evidence of strong aural skills, whatever the achievement, while high achievers who work very hard but do not have strong inherent aural skills did not perform well on this test. How much a factor the medium itself was in performance is difficult to gauge from these results, though the annotated tables indicate some possible inferences.

Table 4: Concurrent Validity Data by Examinee

Examinee	Year	Grades--Written Theory	Grades--Aural Skills	Anecdotal Evaluation
1	jr	A	A	excellent ears
2	sr	C/D	C/D	phenomenal ear but weirdly affected by time of day; also lazy
3	gr			excellent ears
4	soph	B	tested out	excellent ears
5	soph	A	A	excellent ears
6	gr			
7	soph	A	A	excellent ears
8	jr	A/B	A	good ears
9	jr	B	B	very good written work
10	gr			very smart; organist
11	jr	A	tested out	excellent ears
12	gr	B	B	ok ears--not great
13	gr	B	B	hard worker; slow
14	sr	C	C	mediocre ears but very computer literate
15	soph	C	C	Severe language problems; spectacular ears
16	sr	C	C	good ears; lazy
17	jr	C	C/D	very mediocre abilities
18	sr	C	C/D	above average ears
19	sr	C/D	C/D	weak student; mediocre ears
20	jr	B+	B+	good ears; works hard
21	jr	special student		very inconsistent performer
22	sr	A	A	overachiever; better than average ear, but not outstanding
23	soph	C/D	A	excellent ears
24	sr	A	A	excellent ears
25	soph	B	B	slightly above average to mediocre ears
26	jr	B	A	language problems; high ability
27	soph	B+/A-	B+/A-	good ear; perfect pitch
28	jr	B+	A	possible language problems
29	spe			
30	sr	C/D	B+	good ears
31	soph	na	na	good ears; language problems
32	jr	B	B	above average ears
33	gr			ok ears; hard worker
34	fresh	na	na	good ears; hard worker
35	jr	na	na	mediocre ears; hard worker
36	gr	A	A	excellent student; good ears
37	soph	B	B	ok ears
38	soph	B	B	
39	jr	C	C	hard worker; not much natural ability; very computer literate

Table 4: Concurrent Validity Data by Examinee

Examinee	Year	Grades–Written Theory	Grades–Aural Skills	Anecdotal Evaluation
40	gr	B	B	inconsistent: ears ok but perhaps not a lot of native ability
41	jr	B	B	language problems; computer literate
42	jr	C	C	not very good ears
43	soph	B	C	hard worker; not a good ear
44	soph	C	C	not a good student; below average ears
45	soph	B-	B	
46	soph	A	C	excellent written skills; weak ears
47	jr	C	C	much weaker than # 46
48	soph	C	C/D	weak ears; struggles
49	jr	na	na	
50	jr	C	D	lazy
51	gr	C	C-	mediocre
52	jr	C	C-	tries hard; low ability
53	sr	C	C	very computer literate
54	jr	na	na	very weak student
55	gr	C/D	C/D	weak in every way
56	gr	C/D	C/D	hard worker; bad ears
57	sr	C/D	C/D	very bad ears and low ability
58	sr	C/D	C/D	very bad ears and low ability
59	gr	C/D	C/D	very bad ears and low ability
60	soph	C/D	C/D	very bad ears and low ability
61	gr	C/D	C/D	very bad ears and low ability
62	soph	C/D	C/D	very bad ears and low ability
63	spe	na	na	very bad ears and low ability
64	jr	C/D	C/D	very bad ears and low ability
65	jr	C/D	C/D	very bad ears and low ability